

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2005, 02:16:01 ; Search time 554 seconds
(without alignments)
6979.281 Million cell updates/sec

Title: US-10-054-498-1

Perfect score: 2363
Sequence: 1 tcgagccgccttcacaggga.....taataaagatggtttgattc 2363

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

1: Issued Patents NA: *
2: /cgn2_6/ptodata/1/lna/5A_COMB.seq: *
3: /cgn2_6/ptodata/1/lna/5B_COMB.seq: *
4: /cgn2_6/ptodata/1/lna/6A_COMB.seq: *
5: /cgn2_6/ptodata/1/lna/6B_COMB.seq: *
6: /cgn2_6/ptodata/1/lna/PTUS_COMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2363	100.0	2363	3	US-09-742-703-3
2	1593.6	67.4	1783	3	US-09-510-738A-188
3	1593.6	67.4	1783	4	US-09-861-966-188
4	1593.6	67.4	1783	4	US-09-915-048-188
5	1593.6	67.4	1783	4	US-09-949-016-1311
6	1269	53.7	1615	4	US-09-820-002-1
7	1008.2	42.7	1605	2	US-09-000-846-1
8	806.8	34.1	1303.7	4	US-09-949-016-13053
9	391.2	16.6	614	3	US-09-280-116-71
10	325	13.8	21784	4	US-09-820-002-3
11	192	8.1	601	4	US-09-820-002-15
12	192	8.1	601	4	US-09-949-016-45925
13	191	8.1	601	4	US-09-820-002-16
14	169.4	7.2	601	4	US-09-820-002-14
15	169.4	7.2	601	4	US-09-949-016-45924
16	148.4	6.3	1434	4	US-10-177-661-1
17	148.4	6.3	1748	4	US-09-879-792-11
18	145.8	6.2	1341	4	US-10-177-661-3
19	141.4	6.0	2413	3	US-09-518-046-1
20	130	5.5	2440	4	US-09-949-016-5210
21	130	5.5	2440	4	US-09-949-016-5211
22	130	5.5	2440	4	US-09-949-016-5212
23	127.4	5.4	2479	3	US-09-342-749-29
24	127.4	5.4	2479	3	US-09-691-840-29
25	127.4	5.4	2479	3	US-09-685-166A-894
26	127.4	5.4	2479	4	US-09-679-426-894
27	127.4	5.4	2479	4	US-09-759-143-894

ALIGNMENTS

28	124.8	5.3	1077	3	US-08-807-151-2	Sequence 2, Appl1
29	124.8	5.3	1077	3	US-09-478-957-2	Sequence 2, Appl1
30	124.6	5.3	2544	3	US-09-518-046-3	Sequence 3, Appl1
31	124.2	5.3	3245	4	US-09-949-016-5203	Sequence 3203, Ap
32	124.2	5.3	3245	4	US-09-759-143-929	Sequence 929, App
33	123.4	5.2	1479	3	US-09-342-749-1	Sequence 1, Appl1
34	123.4	5.2	1479	3	US-09-691-840-1	Sequence 910, Appl1
35	121.8	5.2	1479	4	US-09-759-143-930	Sequence 931, App
36	120.6	5.1	1476	4	US-09-759-143-931	Sequence 931, App
37	120	5.1	2416	3	US-09-261-416-1	Sequence 256, App
38	117.2	5.0	1100	4	US-09-907-794A-256	Sequence 256, App
39	117.2	5.0	1100	4	US-09-905-125A-256	Sequence 256, App
40	117.2	5.0	1100	4	US-09-902-775A-256	Sequence 256, App
41	117.2	5.0	1100	4	US-09-906-700-256	Sequence 256, App
42	117.2	5.0	1100	4	US-09-903-603A-256	Sequence 256, App
43	117.2	5.0	1100	4	US-09-904-920A-256	Sequence 256, App
44	117.2	5.0	1100	4	US-09-899-064-256	Sequence 256, App
45	117.2	5.0	1100	4	US-09-905-381A-256	Sequence 256, App

RESULT 1	US-09-742-703-3
Sequence 3, Application US/09742703	
Patent No. 6423543	
GENERAL INFORMATION:	
APPLICANT: Lex M. Cowsett	
INVENTOR: Patrick Allen Marcotte	
TITLE OF INVENTION: ANTISENSE MODULATION OF HEPESIN EXPRESSION	
FILE REFERENCE: RTS-0090	
CURRENT FILING DATE: 2000-12-20	
NUMBER OF SEQ ID NOS: 49	
SEQ ID NO 3	
LENGTH: 2363	
TYPE: DNA	
ORGANISM: Homo sapiens	
FEATURE:	
NAME/KEY: CDS	
LOCATION: (826)..(2079)	
US-09-742-703-3	

Query Match	100.0%;	Score 2363;	DB 3;	Length 2363;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2363;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	TCGAGCCCCGCTTTCCAGGAGACCTTAAGAGGGCCACAGGTGAGGACCTTGCCCTAGC	60
DB	1	TCGAGCCCCGCTTTCCAGGAGACCTTAAGAGGGCCACAGGTGAGGACCTTGCCCTAGC	60
QY	61	AGGCCCCCAGCCAGCCGCTTGCTCCAGGCGCCGCTGCGGGGCCACATGCTCC	120
DB	61	AGGCCCCCAGCCAGCCGCTTGCTCCAGGCGCCGCTGCGGGGCCACATGCTCC	120
QY	121	TGCCAGGCTTGAGAGTACCCGAGCCGCGGCACTTGAAGCTTCGCCCTGAC	180
DB	121	TGCCAGGCTTGAGAGTACCCGAGCCGCGGCACTTGAAGCTTCGCCCTGAC	180
QY	181	TGAGCCCAAGGTAAGAGCAAGGGCCCCAGACTCAAGTTCAGCCCTGAGCAGGG	240
DB	181	TGAGCCCAAGGTAAGAGCAAGGGCCCCAGACTCAAGTTCAGCCCTGAGCAGGG	240
QY	241	TTCCCTCATCCCCCAGCCAGCTTAATGCCCACCTCTTAATAGAGGGTTCTGGGAGC	300
DB	241	TTCCCTCATCCCCCAGCCAGCTTAATGCCCACCTCTTAATAGAGGGTTCTGGGAGC	300
QY	301	TGAAGAGGGGCACTATGAGCTTCCCAAGACCTTAGGTGTTCTGCTGCTCTT	360
DB	301	TGAAGAGGGGCACTATGAGCTTCCCAAGACCTTAGGTGTTCTGCTGCTCTT	360
QY	361	CAGACTCAGCGCTTGAGACCCCAAGTCTTCTCCCAAGACCCAGAGACTTCAGCCCTCAG	420

Db 361 CAGACTAGCGGCTGGAGCCCAAGTCTTCTCCCAAGACCCAGAGTTCAGGCCCTCAG 420
Qy 421 GCCCTCTCTCTCTATCTAGAGAGTCTGGCCCCCAATTCCTCTTCCCAAGACTTA 480
Db 421 GCCCTCTCTCTCTATCTAGAGAGTCTGGCCCCCAATTCCTCTTCCCAAGACTTA 480
Qy 481 TGATTTAGAGTCTCAGCTGTCTCTCTCCCAACCCGGGATCTCAGTCCCTCTCTCAG 540
Db 481 TGATTTAGAGTCTCAGCTGTCTCTCTCCCAACCCGGGATCTCAGTCCCTCTCTCAG 540
Qy 541 CAGGCTCAGGAGTGGGGTCCCAATCCCTGCAATCCAGAGGCTCCCGCTGTGTCTCA 600
Db 541 CAGGCTCAGGAGTGGGGTCCCAATCCCTGCAATCCAGAGGCTCCCGCTGTGTCTCA 600
Qy 601 GACACTGACCCCATCTTGAACCCAGCCCAATCTGCTCTCTGATCAAGGCTCTCTG 660
Db 601 GACACTGACCCCATCTTGAACCCAGCCCAATCTGCTCTCTGATCAAGGCTCTCTG 660
Qy 661 CCAAGGCTCAGTCTCTCAGCTCTGCTGATGAGAGGCTGGGAGCTGGGGGCTGAGACT 720
Db 661 CCAAGGCTCAGTCTCTCAGCTCTGCTGATGAGAGGCTGGGAGCTGGGGGCTGAGACT 720
Qy 721 GGGCTGGGCTGGGCTCCCAAGGCTCTGCTCTCCCTCTCTCTCTCTCTCTCTCTCT 780
Db 721 GGGCTGGGCTGGGCTCCCAAGGCTCTGCTCTCCCTCTCTCTCTCTCTCTCTCTCT 780
Qy 781 TGCGCCAGAGAGTCAAGCCAGGAATCAATTAACAAGAGCAGTGAATGCGCCAGAGAG 840
Db 781 TGCGCCAGAGAGTCAAGCCAGGAATCAATTAACAAGAGCAGTGAATGCGCCAGAGAG 840
Qy 841 GGTGGCCGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 841 GGTGGCCGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Qy 901 CTGCTACTTCTGACAGCCATGCGGAGCGGCACTCTGCGGCAATGTGTCTCTCTCT 960
Db 901 CTGCTACTTCTGACAGCCATGCGGAGCGGCACTCTGCGGCAATGTGTCTCTCTCT 960
Qy 961 AGTGAACAGAGAGCGGCTGTATCCCAAGTGAAGTCAAGTCTGCGGCTCTCTCTCT 1020
Db 961 AGTGAACAGAGAGCGGCTGTATCCCAAGTGAAGTCAAGTCTGCGGCTCTCTCTCT 1020
Qy 1021 TTTGACAAGAGAGAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db 1021 TTTGACAAGAGAGAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Qy 1081 GCCGAGCTCAGCTGCGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 1081 GCCGAGCTCAGCTGCGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Qy 1141 GTGGAACAGAGAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 1141 GTGGAACAGAGAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Qy 1201 CCCCACACCCAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db 1201 CCCCACACCCAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Qy 1261 TTGGCCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 1261 TTGGCCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Qy 1321 GGGCCGAGACACAGCTGTGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Db 1321 GGGCCGAGACACAGCTGTGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Qy 1381 CACTCTGTGGGAGTCT 1440
Db 1381 CACTCTGTGGGAGTCT 1440
Qy 1441 CCGAGGCGAACCAGGCT 1500
Db 1441 CCGAGGCGAACCAGGCT 1500

Db 1441 CCGAGGCGAACCAGGCT 1500
Qy 1501 TCTCCCAAGGCTCTGAGCTGTGGGCTGAGGCTGTGCTCTCAACAGGGGGCTATCTCTCC 1560
Db 1501 TCTCCCAAGGCTCTGAGCTGTGGGCTGAGGCTGTGCTCTCAACAGGGGGCTATCTCTCC 1560
Qy 1561 TTTTGGGAGCCCAACAGCGAG 1620
Db 1561 TTTTGGGAGCCCAACAGCGAG 1620
Qy 1621 CCCCTGCTCTCAAGATATCAATCAGCTGTGTGCTCTCCAGTGTGCTGCTGCTGCTGCTG 1680
Db 1621 CCCCTGCTCTCAAGATATCAATCAGCTGTGTGCTCTCCAGTGTGCTGCTGCTGCTGCTG 1680
Qy 1681 GTGGATGGCAAGATCTGTATCCGTAAGGAGGCTGGGGGCAACGAGTACTATGGCAACAG 1740
Db 1681 GTGGATGGCAAGATCTGTATCCGTAAGGAGGCTGGGGGCAACGAGTACTATGGCAACAG 1740
Qy 1741 GCCGGGCTACTCAGAGAGGCTGAGTCCCAATATCAGCAATGATGTCTGCAATGGGCT 1800
Db 1741 GCCGGGCTACTCAGAGAGGCTGAGTCCCAATATCAGCAATGATGTCTGCAATGGGCT 1800
Qy 1801 GACTTCTATGAAACAGATCAAGCCCAAGATGTCTGTCTGCTAACCAGAGGCTGCT 1860
Db 1801 GACTTCTATGAAACAGATCAAGCCCAAGATGTCTGTCTGCTAACCAGAGGCTGCT 1860
Qy 1861 ATTGAATGCTGCAAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1920
Db 1861 ATTGAATGCTGCAAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1920
Qy 1921 ACCGCACTGTGGGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
Db 1921 ACCGCACTGTGGGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
Qy 1981 AAGCAGGCGCTCTACCAAAAGTCAAGTCTTCCGGAGTGTGATCTTCCAGGCTATTAAG 2040
Db 1981 AAGCAGGCGCTCTACCAAAAGTCAAGTCTTCCGGAGTGTGATCTTCCAGGCTATTAAG 2040
Qy 2041 ACTCACTCGAAGCAGCGGCTGAGTGAACCAAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 2100
Db 2041 ACTCACTCGAAGCAGCGGCTGAGTGAACCAAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 2100
Qy 2101 GCTCTCAGGCGCGAGTGTATCCGCTGTGTGAGATCCAGCTGTGCGAGAGATGGAGCT 2160
Db 2101 GCTCTCAGGCGCGAGTGTATCCGCTGTGTGAGATCCAGCTGTGCGAGAGATGGAGCT 2160
Qy 2161 TTTTCTTCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
Db 2161 TTTTCTTCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
Qy 2221 CACAGTGGGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280
Db 2221 CACAGTGGGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280
Qy 2281 AAATATTTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340
Db 2281 AAATATTTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340
Qy 2341 AAATATTAAGATGTGTTTGAAT 2363
Db 2341 AAATATTAAGATGTGTTTGAAT 2363

RESULT 2
US-09-510-738A-188
Sequence 188, Application US/09510738A
Patent No. 6268165
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
Ovarian Cancer
FILE REFERENCE: D6223CIP-A
CURRENT APPLICATION NUMBER: US/09/510, 738A

CURRENT FILING DATE: 2000-02-22
 PRIOR APPLICATION NUMBER: 09/039,211
 PRIOR FILING DATE: 03-14-1998
 NUMBER OF SEQ ID NOS: 188
 SEQ ID NO 188
 LENGTH: 1783
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: full length cDNA of hepsin
 US-09-510-738A-188

Query Match 67.4%; Score 1593.6; DB 3; Length 1783;
 Best Local Similarity 99.1%; Pred. No. 0;
 Matches 1513; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

736 CCCCCAGGCGCTGCTCCCGCTCAGTCTCTCAGAGTCCACCTGCGCCAGAGATCA 795
 157 CCGAGAGGCTCCGCCCCCAGCTGAGACCCCA -GGGTCCACCTTGCCCAAGAGTCA 215
 796 GCCAGGAATCATTTAAACAAGAGCAGTGAATGCGCGCAAGAGAGGTTGGAGCTGTG 855
 216 GCCAGGGAATCATTTAAACAAGAGCAGTGAATGCGCGCAAGAGAGGTTGGAGCTGTG 275
 856 CCATGCTCTCCAGACCCCAAGTGGACCTCTCACTGGGAGACCTTGCTACTTTGACA 915
 276 CCATGCTCTCCAGACCCCAAGTGGACCTCTCACTGGGAGACCTTGCTACTTTGACA 335
 916 GCCATCGGGGCGGCACTCTGCGGCAATGTTGCTTTCTCTCAGAGTGAACAGAGGCG 975
 336 GCCATCGGGGCGGCACTCTGCGGCAATGTTGCTTTCTCTCAGAGTGAACAGAGGCG 395
 976 CTGTACCAAGTGCAGTCACTCTGCGGAGCGCTCGGCTCATGTGTTTGAACAAGCGAA 1035
 396 CTGTACCAAGTGCAGTCACTCTGCGGAGCGCTCGGCTCATGTGTTTGAACAAGCGAA 455
 1036 GGGAGCTGGCGGCTGCTGCTCTCGGCTCCCAAGCGAGGTAGCCGGAAGTCACTGTC 1095
 456 GGGAGCTGGCGGCTGCTGCTCTCGGCTCCCAAGCGAGGTAGCCGGAAGTCACTGTC 515
 1096 GAGAGATGAGGCTTCTCAGAGGCACTGACCCACTCCGAGTGAAGTGAACGAGCGGCG 1155
 516 GAGAGATGAGGCTTCTCAGAGGCACTGACCCACTCCGAGTGAAGTGAACGAGCGGCG 575
 1156 GCCAATGAGCAGTGGGCTTTCTTGTGTGAGCAAGAGGAGGCTGCCCAACCCAGAGG 1215
 576 GCCAATGAGCAGTGGGCTTTCTTGTGTGAGCAAGAGGAGGCTGCCCAACCCAGAGG 635
 1216 CTGCTGAGAGTCACTCTCGGTGTGATGTCCTCCAGAGCGCTTTTGGCCGCACTGTC 1275
 636 CTGCTGAGAGTCACTCTCGGTGTGATGTCCTCCAGAGCGCTTTTGGCCGCACTGTC 695
 1276 CAAGACTGAGCCGAGAGAGCTGCGCTGAGACCGCATCTGTGGAGGCGGAGACACAGC 1335
 696 CAAGACTGAGCCGAGAGAGCTGCGCTGAGACCGCATCTGTGGAGGCGGAGACACAGC 755
 1336 TTGAGCCGAGTGGCGTGGCAAGTCAAGCTTGTGATGATGAGCAACACTTGTGGAGGA 1395
 756 TTGAGCCGAGTGGCGTGGCAAGTCAAGCTTGTGATGATGAGCAACACTTGTGGAGGA 815
 1396 TCCCTGCTCTCCGAGGAGTGGGTGTGACAGCCGCGCACTGCTTCCGAGAGCGAATCGG 1455
 816 TCCCTGCTCTCCGAGGAGTGGGTGTGACAGCCGCGCACTGCTTCCGAGAGCGAATCGG 875
 1456 GTCTGCTCCGAGTGGCGAGTGTGTCGCGGTGGCGGAGCGGCTTCCCAAGGAGTGTG 1515
 876 GTCTGCTCCGAGTGGCGAGTGTGTCGCGGTGGCGGAGCGGCTTCCCAAGGAGTGTG 935
 1516 CAGCTGGGAGTGCAGAGCTGTGTCTAACACGAGGAGCTTCTTCCCTTTCGAGACCCCAAC 1575
 936 CAGCTGGGAGTGCAGAGCTGTGTCTAACACGAGGAGCTTCTTCCCTTTCGAGACCCCAAC 995
 1576 AGCAGAGGAACAGACGATATTGCTGCTGCTCACTCTCAAGTCCCTGCTCCACA 1635

996 AGCAGAGGAACAGACGATATTGCTGCTGCTCACTCTCAAGTCCCTGCTCCACA 1055
 1636 GAATACATCCAGCTGTGTGCTCCAGCTCCGAGCGGCGCTGTGATGAGCAAGATC 1695
 1056 GAATACATCCAGCTGTGTGCTCCAGCTCCGAGCGGCGCTGTGATGAGCAAGATC 1115
 1696 TGTACCTGAGCGGCTGGGCGCAACGAGATCTATGAGCAACAGGCGGGGTACTCCAG 1755
 1116 TGTACCTGAGCGGCTGGGCGCAACGAGATCTATGAGCAACAGGCGGGGTACTCCAG 1175
 1756 GAGGCTGAGTCCCATATATGAGCAATGATGTCTGCAATGAGGCGGCTGATGATGAAC 1815
 1176 GAGGCTGAGTCCCATATATGAGCAATGATGTCTGCAATGAGGCGGCTGATGATGAAC 1235
 1816 CAGATCAAGCCCAAGATGTTCTGTGTGCTGCTACCCGAGGAGTGGCATTTGATGCTCCAG 1875
 1236 CAGATCAAGCCCAAGATGTTCTGTGTGCTGCTACCCGAGGAGTGGCATTTGATGCTCCAG 1295
 1876 GCGGACAGCGGCTGCTTCTTGTGTGAGGACAGCATCTTCCGAGCCCACTGTTGGCG 1935
 1296 GCGGACAGCGGCTGCTTCTTGTGTGAGGACAGCATCTTCCGAGCCCACTGTTGGCG 1355
 1936 CTGTGTGCAATTTGAGTGTGGGCACTGGGCTGTGCTGCTGCGCCAGAGCCAGGCGTCTAC 1995
 1356 CTGTGTGCAATTTGAGTGTGGGCACTGGGCTGTGCTGCTGCGCCAGAGCCAGGCGTCTAC 1415
 1996 ACCAAGTCAAGTGAATTTCCGAGAGTGAATCTTCCAGAGCCATTAAGACTCACTCCGAGCC 2055
 1416 ACCAAGTCAAGTGAATTTCCGAGAGTGAATCTTCCAGAGCCATTAAGACTCACTCCGAGCC 1475
 2056 AGCGGATGTGACCCAGCTTGAACCGGTGCTTCTGCTGCGCAAGCTTCCAGAGCCGGA 2115
 1476 AGCGGATGTGACCCAGCTTGAACCGGTGCTTCTGCTGCGCAAGCTTCCAGAGCCGGA 1535
 2116 GGTATCCCGGTGTGGGATCCAGGCTGGGCGAGAGTGGAGCGTTTCTTCTTGGGCG 2175
 1536 GGTATCCCGGTGTGGGATCCAGGCTGGGCGAGAGTGGAGCGTTTCTTCTTGGGCG 1595
 2176 CGGTCCACAGTCCCAAGGACACCTTCCCTCCAGGCTCTCTTCCACAGTGGCGGCGC 2235
 1596 CGGTCCACAGTCCCAAGGACACCTTCCCTCCAGGCTCTCTTCCACAGTGGCGGCGC 1655
 2236 ACTAGCCCGGAGACCAACCACTCACTCTGAGCCCATGTAATTTGTTGCT 2295
 1656 ACTAGCCCGGAGACCAACCACTCACTCTGAGCCCATGTAATTTGTTGCT 1715
 2296 GTCTGAGAGTCTCTGATGAGTGGGCGGCTGATGATGAGTGGTCTTAAATTAAGATGG 2355
 1716 GTCTGAGAGTCTCTGATGAGTGGGCGGCTGATGATGAGTGGTCTTAAATTAAGATGG 1775
 2356 TTTTGATT 2363
 1776 TTTTGATT 1783

RESULT 3
 US-09-861-966-188
 Sequence 188, Application US/09661966
 Patent No. 6518028
 GENERAL INFORMATION:
 APPLICANT: O'Brien, Timothy J.
 TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
 FILE REFERENCE: D6223CIP-A/Div
 CURRENT APPLICATION NUMBER: US/09/861,966
 PRIOR FILING DATE: 2001-05-21
 PRIOR APPLICATION NUMBER: 09/510,738
 NUMBER OF SEQ ID NOS: 188
 SEQ ID NO 188
 LENGTH: 1783
 TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: full length cDNA of hepsin
US-09-861-966-188

Query Match 67.4%; Score 1593.6; DB 4; Length 1783;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 736 CCCCCAGGCTCCCTCCCTCCATCTCTCAAGAGTCCACCTGGGCCAGAGAGTCA 795
DB 157 CTTGAGAGCTCCGCCCCCAGCTCTGAGACCCCA-GGGTCCCACTGGCCAGAGGTC 215
QY 796 GCCAGGGAGATTAACAAGAGAGAGTACATGGGCGGAGAGAGGGGGCGGACTGTG 855
DB 216 GCCAGGGAGATTAACAAGAGAGAGTACATGGGCGGAGAGAGGGGGCGGACTGTG 275
QY 856 CCATGCTGCTCCAGACCCAGAGTGGAGCTCTCACTGGGAGACCTGCTACTTCTGACA 915
DB 276 CCATGCTGCTCCAGACCCAGAGTGGAGCTCTCACTGGGAGACCTGCTACTTCTGACA 335
QY 916 GCCATCGGGGCGGATCTCTGGGCAATTGGCTGTTCTCTCAAGAGTACACAGAGCCG 975
DB 336 GCCATCGGGGCGGATCTCTGGGCAATTGGCTGTTCTCTCAAGAGTACACAGAGCCG 395
QY 976 CTGTACCCAGTGAAGTCAAGCTCTGCGGAGCTGGGCTCATGGTCTTTGAACAAGGAA 1035
DB 396 CTGTACCCAGTGAAGTCAAGCTCTGCGGAGCTGGGCTCATGGTCTTTGAACAAGGAA 455
QY 1036 GGGAGCTGGCGGCTGTGTCTCTCGGCTCAACCGCAGAGTACCGGACTCAGCTGC 1095
DB 456 GGGAGCTGGCGGCTGTGTCTCTCGGCTCAACCGCAGAGTACCGGACTCAGCTGC 515
QY 1096 GAGAGATGGGCTTCTCAAGGCACTGACCTCCAGCTGAGACGTGCAACGGCGGAC 1155
DB 516 GAGAGATGGGCTTCTCAAGGCACTGACCTCCAGCTGAGACGTGCAACGGCGGAC 575
QY 1156 GCCAATGGCAAGTGGGCTTCTGATGAGACGAGGGAGGAGGCGCCCAACCGAGAG 1215
DB 576 GCCAATGGCAAGTGGGCTTCTGATGAGACGAGGGAGGAGGCGCCCAACCGAGAG 635
QY 1216 CTGTGAGAGTCACTCCGATGATGATGACCCAGAGGCGCTTTCTTGCGCCATCTGC 1275
DB 636 CTGTGAGAGTCACTCCGATGATGATGATGACCCAGAGGCGCTTTCTTGCGCCATCTGC 695
QY 1276 CAAGACTGTGGCCGAGAGACTGCGGTGACCGCATCTGTGGAGGCGCGGACACAC 1335
DB 696 CAAGACTGTGGCCGAGAGACTGCGGTGACCGCATCTGTGGAGGCGCGGACACAC 755
QY 1336 TTGGAGCGGATGGCGGTGCAAGCTTGTATGATGAGACACCTCTGTGGGGA 1395
DB 756 TTGGAGCGGATGGCGGTGCAAGCTTGTATGATGAGACACCTCTGTGGGGA 815
QY 1396 TCCCTGCTCTCCGAGGAGCTGGGTCTGACAGCCGCCCACTTCCCGAGAGGAGCCG 1455
DB 816 TCCCTGCTCTCCGAGGAGCTGGGTCTGACAGCCGCCCACTTCCCGAGAGGAGCCG 875
QY 1456 GTCTGTCTCCGATGGAGATTTGGCGGTGCGGTGAGGCGGCTCTCCCAAGGCTG 1515
DB 876 GTCTGTCTCCGATGGAGATTTGGCGGTGCGGTGAGGCGGCTCTCCCAAGGCTG 935
QY 1516 CAGCTGGAGGAGGAGGCTGTGTACACAGGGGAGTATCTTCCCTTTGGGAGCCCAAC 1575
DB 936 CAGCTGGAGGAGGAGGCTGTGTGTACACAGGGGAGTATCTTCCCTTTGGGAGCCCAAC 995
QY 1576 AGCAGAGGAGACAGCAAGATTTGCCCTGTCTCACTTCCAGTCCCTGCTCACA 1635
DB 996 AGCAGAGGAGACAGCAAGATTTGCCCTGTCTCACTTCCAGTCCCTGCTCACA 1055
QY 1636 GAATACATCCAGCTGTGTGCTCCAGACTCCGGCCAGGCGCTGTGTGATGAGCAAGTC 1695
DB 1056 GAATACATCCAGCTGTGTGCTCCAGACTCCGGCCAGGCGCTGTGTGATGAGCAAGTC 1115

QY 1696 TGTACGTGACGGGCTGGGCAACAGCAGTACTATGGCCAAAGGCGGGGTACTCAG 1755
DB 1116 TGTACGTGACGGGCTGGGCAACAGCAGTACTATGGCCAAAGGCGGGGTACTCAG 1175
QY 1756 GAGGCTGAGTCCCATTAATCAGCAATGATGTCTGCATGAGCGCTGACTTCTATGAAAC 1815
DB 1176 GAGGCTGAGTCCCATTAATCAGCAATGATGTCTGCATGAGCGCTGACTTCTATGAAAC 1235
QY 1816 CAGATCAAGCCAAAGATTTCTGTGCTGCTACCCGAGGGGTGCAATTGATGCTGCAG 1875
DB 1236 CAGATCAAGCCAAAGATTTCTGTGCTGCTACCCGAGGGGTGCAATTGATGCTGCAG 1295
QY 1876 GCGACAGCGGTGCTCTTGTGTGTGAGAGACAGACTCTCTGAGCGCCACGTTGGCG 1935
DB 1296 GCGACAGCGGTGCTCTTGTGTGTGAGAGACAGACTCTCTGAGCGCCACGTTGGCG 1355
QY 1936 CTGTGTGCAATTGATGAGTGGGCACTGCTGTGCTGCTGAGAGAGAGAGAGAGAG 1995
DB 1356 CTGTGTGCAATTGATGAGTGGGCACTGCTGTGCTGCTGAGAGAGAGAGAGAGAG 1415
QY 1996 ACCAATGCAAGTCTCCGGGAGTGGATCTTCAGGSCATTAAGACTCATCCTCGAAGCC 2055
DB 1416 ACCAATGCAAGTCTCCGGGAGTGGATCTTCAGGSCATTAAGACTCATCCTCGAAGCC 1475
QY 2056 AGCGCATGGTGAACCCAGCTCTGACCGGTGCTTCTGCTGCGAGGCTCCAGGCGCGA 2115
DB 1476 AGCGCATGGTGAACCCAGCTCTGACCGGTGCTTCTGCTGCGAGCTCCAGGCGCGA 1535
QY 2116 GGTGATCCCGGTGAGTGAATCCAGCTGGGCCAGAGATGGAGAGTTTCTTCTTGGGCC 2175
DB 1536 GGTGATCCCGGTGAGTGAATCCAGCTGGGCCAGAGATGGAGAGTTTCTTCTTGGGCC 1595
QY 2176 CGGTCCACAGTCCAGAGACACCTCCCTCAGGGTCTCTTCCACAGTGGCGGCGC 2235
DB 1596 CGGTCCACAGTCCAGAGACACCTCCCTCAGGGTCTCTTCCACAGTGGCGGCGC 1655
QY 2236 ACTAGGCCCGGAGACCAACCACTCACTCTCTGACCCCAATGTAATATTGTCTGCT 2295
DB 1656 ACTAGGCCCGGAGACCAACCACTCACTCTCTGACCCCAATGTAATATTGTCTGCT 1715
QY 2296 GTCTGGAGCTCTGTCTAGGTGCTGCTGATGATGAGTCTCTTTAATTAATAAGTGG 2355
DB 1716 GTCTGGAGCTCTGTCTAGGTGCTGCTGATGATGAGTCTCTTTAATTAATAAGTGG 1775
QY 2356 TTTTGATT 2363
DB 1776 TTTTGATT 1783

RESULT 4
US-09-919-048-188
Sequence 188, Application US/09919048
Patent No. 6787354
GENERAL INFORMATION:
APPLICANT: Timothy J.
APPLICANT: O'Brien, Martin J.
APPLICANT: Santini, Alessandro
TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer
FILE REFERENCE: D62231P/A/D/CIP
CURRENT APPLICATION NUMBER: US/09/919,048
PRIOR APPLICATION NUMBER: 2001-07-30
PRIOR FILING DATE: 2001-05-21
NUMBER OF SEQ ID NOS: 188
SEQ ID NO 188
LENGTH: 1783
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: full length cDNA of hepsin
US-09-919-048-188
Query Match 67.4%; Score 1593.6; DB 4; Length 1783;

QY 935 GGGCCATTGTGGCTTTTCTCTCAGAGAGTACAGAGCCGCTGTAACCAAGTACAGTCA 994
DB 287 GGGCCATTGTGGCTTTTCTCTCAGAGAGTACAGAGCCGCTGTAACCAAGTACAGTCA 346
QY 995 GCTCTGCGGACGCTGCGCTCATGTCTTTGACAAAGAGGAGAGTGGCGGCTGCT 1054
DB 347 GCTCTGCGGACGCTGCGCTCATGTCTTTGACAAAGAGGAGAGTGGCGGCTGCT 406
QY 1055 GCTCTGCGGCTGCGGCTCATGTCTTTGACAAAGAGGAGAGTGGCGGCTGCT 1114
DB 407 GCTCTGCGGCTGCGGCTCATGTCTTTGACAAAGAGGAGAGTGGCGGCTGCT 466
QY 1115 GGGCACTAGCCCACTCCAGAGTGGAGAGTGGAGAGGCGGCGCAATGACAGTGGGCT 1174
DB 467 ----- 466
QY 1175 TCTTGTGTGTGAGAGAGGAGAGGCTGCCACACCCAGAGGCTGTGAGAGTCACTCCG 1234
DB 467 ----- 466
QY 1235 TGTGTGATTTGCCCCAGAGGCGGCTTTCTTGGCCGCAATCTGCCAGAGTGTGGCCGAGAGA 1294
DB 467 ---GTGATTTGCCCCAGAGGCGGCTTTCTTGGCCGCAATCTGCCAGAGTGTGGCCGAGAGA 523
QY 1295 AGCTGCCCCGTGAGACCGCATCTGTGGAGAGCCGAGACACCAAGCTTGGCGGCTGGCCGCTGGC 1354
DB 524 AGCTGCCCCGTGAGACCGCATCTGTGGAGAGCCGAGACACCAAGCTTGGCGGCTGGCCGCTGGC 583
QY 1355 AAGTCAGGCTTGCATGATGAGAGACACTCTGTGGAGAGTCCCTGCTCTCCGAGGAGCT 1414
DB 584 AAGTCAGGCTTGCATGATGAGAGACACTCTGTGGAGAGTCCCTGCTCTCCGAGGAGCT 643
QY 1415 GGGTGTCTGACAGCGCCGCACTGCTTCCGAGAGCGGAGTCTGTCCGAGTGGCGAG 1474
DB 644 GGGTGTCTGACAGCGCCGCACTGCTTCCGAGAGCGGAGTCTGTCCGAGTGGCGAG 703
QY 1475 TGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 1534
DB 704 TGTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 763
QY 1535 TGTGTGACAGCGGAGGCTATCTCTCTTCCGAGAGCCCAAGAGGAGAGAGAGAGAGAGAG 1594
DB 764 TGTGTGACAGCGGAGGCTATCTCTCTTCCGAGAGCCCAAGAGGAGAGAGAGAGAGAGAG 823
QY 1595 ATATTGCCCTGTGCTCACTCTTCAAGTCCCTGCGCCCTCAAGAGATATCAAGCTGTGT 1654
DB 824 ATATTGCCCTGTGCTCACTCTTCAAGTCCCTGCGCCCTCAAGAGATATCAAGCTGTGT 883
QY 1655 GCTTCCCGAGCTGCGGCGGAGGCTGTGTGATGAGAGAGTGTGACAGGCTGGG 1714
DB 884 GCTTCCCGAGCTGCGGCGGAGGCTGTGTGATGAGAGAGTGTGACAGGCTGGG 943
QY 1715 GCAACAGAGAGTACTAGTGGAGAGAGGCGGAGGCTCAAGAGAGGCTGAGTCCCATTA 1774
DB 944 GCAACAGAGAGTACTAGTGGAGAGAGGCGGAGGCTCAAGAGAGGCTGAGTCCCATTA 1003
QY 1775 TCAGCAATGATGTCTGCAATGAGCGCTGACTTCTATGAAACAGATCAAGCCAGAGTGT 1834
DB 1004 TCAGCAATGATGTCTGCAATGAGCGCTGACTTCTATGAAACAGATCAAGCCAGAGTGT 1063
QY 1835 TCTGTGTGCTGCTACCCCGAGAGGCTGAGTCTTCCAGAGGAGAGAGCGGCTGCTCC 1894
DB 1064 TCTGTGTGCTGCTACCCCGAGAGGCTGAGTCTTCCAGAGGAGAGAGCGGCTGCTCC 1123
QY 1895 TTTGTGTGAGAGAGAGAGTCTCTGAGAGAGAGAGTGTGAGAGTGTGAGAGTGT 1954
DB 1124 TTTGTGTGAGAGAGAGAGTCTCTGAGAGAGAGAGTGTGAGAGTGTGAGAGTGT 1183
QY 1955 GGGGCACTGTGCTGCTGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2014
DB 1184 GGGGCACTGTGCTGCTGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1243

QY 2015 GGGAGTGATCTTCAGGCGCATTAAGACTCACTCCGAAGCCAGCGGAGTGTGAGCCAGC 2074
DB 1244 GGGAGTGATCTTCAGGCGCATTAAGACTCACTCCGAAGCCAGCGGAGTGTGAGCCAGC 1303
QY 2075 TCTGACCGGTGAGCTTCTGCTGCGGAGCGCTCAGAGGCGCGAGTGTGAGTGTGAG 2134
DB 1304 TCTGACCGGTGAGCTTCTGCTGCGGAGCGCTCAGAGGCGCGAGTGTGAGTGTGAG 1363
QY 2135 TCTGACCGGTGAGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2194
DB 1364 TCTGACCGGTGAGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1423
QY 2195 CACCTCCCTCAGAGGCTCTCTCTTCCAGAGTGTGAGGCGGCGCACTGAGCCGAGAGCAC 2254
DB 1424 CACCTCCCTCAGAGGCTCTCTCTTCCAGAGTGTGAGGCGGCGCACTGAGCCGAGAGCAC 1483
QY 2255 CAACCTCAGCTCTGAGCCCGGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2314
DB 1484 CAACCTCAGCTCTGAGCCCGGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1543
QY 2315 GTGCCCTGATGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2363
DB 1544 GTGCCCTGATGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1592

RESULT 7
US-09-000-846-1
Sequence 1, Application US/09000846
Patent No. 5981830
GENERAL INFORMATION:
APPLICANT: WU, QINGYU
APPLICANT: SADLER, JASPER
TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: MILLER, WHITE, ZELANO & BRANIGAN, P. C.
STREET: 2200 CLARENDON BLVD, SUITE 1400
CITY: ARLINGTON
STATE: VA
COUNTRY: US
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,846
FILING DATE: 30-DEC-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/866,058
FILING DATE: 30-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: LEBOVITZ, RICHARD M.
REGISTRATION NUMBER: 37,067
REFERENCE/DOCKET NUMBER: BERLX 65P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-243-6333
TELEFAX: 703-243-6410
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1605 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: linear
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:

```

NAME/KEY: CDS
LOCATION: 35..1282
US-09-000-846-1

Query Match      42.7%; Score 1008.2; DB 2; Length 1605;
Best Local Similarity 80.2%; Pred. No. 7e-243;
Matches 1283; Conservative 0; Mismatches 278; Indels 38; Gaps 7;

OY 792 CTCAGCCGAGGATATTATTAACAAGGCACTGACATGCGCCAGAAAGAGGCTGGCCGAC 851
DB 1 GTCAACCTGGGAAATATTAAACAAGAGTCCCTGACATGCG--GAAGAGGGGTGGCCGAC 57
OY 852 TGTGCATGCTGCTCTCAGACCCCAAGATGAGAGCTCTCACTGCGGGAGACCTGCTACTTCT 911
DB 58 TGCAGATGCTGCTCTCAGACCCCAAGATGAGAGCTCTCACTGCGGGAGACCTGCTACTTCT 117
OY 912 GACAGCCATCGGGGGGAGGATCTGCGGCAATGCTGCTGCTCTCTCAGAGAGTACACAGA 971
DB 118 GACAGCCATGCGGGGGGAGGATCTGCGGCAATGCTGCTGCTCTCTCAGAGAGTACACAGA 177
OY 972 GCCGCTGATACCAAGTGCAGAGTCACTGCTGCGGACGCTGCTCATGCTTTTGAACAAAC 1031
DB 178 GCCATGTAAACCAAGTGCAGAGTCACTGCTGCGGACGCTGCTGCGGCTTTTGAACAAAC 237
OY 1032 GAAAGGAGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1091
DB 238 GAAAGGAGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297
OY 1092 CTGCGAGAGATGGGCTTCTCTCAAGGCACTGACCCACTCCGAGCTGAGCTGCGAAGCGC 1151
DB 298 CTGTAGAGAGATGGGCTTCTCTCAAGGCACTGACCCACTCCGAGCTGAGCTGCGAAGCGC 357
OY 1152 GGGCGCCAAATGAGCACTGGGGCTTCTCTGTGTGAGACGAGGGAGAGGCTGCCACACCA 1211
DB 358 GGGCGCCAAATGAGCACTGGGGCTTCTCTGTGTGAGACGAGGGAGAGGCTGCCACACCA 417
OY 1212 GAGGCTGTGAGAGTCACTCTCGTGTGTGATGATGCCCAAGAGCCGTTCTTGGCCGCCAT 1271
DB 418 GAGGCTGTGAGTCACTCTCGTGTGTGATGATGCTGTGAGAGGCGCATTCCTGACGCGAC 477
OY 1272 CTGCGAAGACTGTGAGCCGAGAGAGCTCCCGTGGACCGCATCTGTGGAGGCGCGGACAC 1331
DB 478 CTGCGAAGACTGTGAGCCGAGAGAGCTCCCGTGGACCGCATCTGTGGAGGCGCGGACAG 537
OY 1332 CAGCTTGGGGCGGCTGGCCGTGGCAAGTCAAGCTTGGCTATGATGAGACACACTTGTGG 1391
DB 538 CAGCTTGGGAAAGTGGCCGTGGCAAGTCAAGCTTGGCTATGATGAGAGACCACTTGTGG 597
OY 1392 GGGATCCCTGCTCTCGGGGAGACTGGGTGTGACAGCCGCCCACTGCTTCCGAGAGCGGAA 1451
DB 598 GGGATCCCTGCTCTCGGGGAGACTGGGTGTGACAGCTGCTGCACTTCTTCCAGAGCGGAA 657
OY 1452 CCGGGTCTGCTCCCGATGGCGAGTGTGCGGGTCCGCTGGCCAGGCTTCTCCGACGG 1511
DB 658 CCGGGTCTGCTCTCGGGTGGCGAGTGTGCTGTGCTGTGAGCCCGGACCTTCAACCCATG 717
OY 1512 TCTGTAGTGGGGGTGAGGCTGTGTGTCTACACAGGGGGCTATCTTCCCTTTTGGAGCC 1571
DB 718 TGTGTAACTGGGGGTGAGGCTGTGTGTATCTATCATGGGGCTATCTTCCCTTTTGAAGCC 777
OY 1572 CAACGAGAGAGAAACAGCAAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1631
DB 778 TACTATCAAGAAACAGCAAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 837
OY 1632 CACAGAAATATCCAGCTGTGTGTGCTTCCAGCTGCGGCGCAAGGCTGTGTGATGGCAA 1691
DB 838 CACAGAAATATCCAGCTGTGTGTGCTTCCAGCTGCGGCGCAAGGCTGTGTGATGGCAA 897
OY 1692 GATCTGTACCTGTGAGGGGCTGGGGCAACAGAGATATATGAGCCAAAGGCGGGGTACT 1751
DB 898 GATCTGTACTGTGAGCGGGCTGGGGTAAACAACAGTTCATGAGCCAAAGGCTGTGTGCT 957
OY 1752 CCAGGAGGCTGAGTCCCATTAATCAGCAATGATGTCTGCAATGAGCGCTGACTTATG 1811

```

```

DB 958 CCAAGAGGCCGGGTTCCCATTAACAGAAAGTTTGAACAAGCCCGACCTTCAAGC 1017
OY 1812 AAACCAATCAAGCCCAAGATTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1871
DB 1018 GAATCAGATCAAGCCCAAGATTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077
OY 1872 CCAGGGGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1931
DB 1078 CCAGGGGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1137
OY 1932 GCGGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1991
DB 1138 GCGGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
OY 1992 CTACACCAAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2051
DB 1198 GTACACCAAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257
OY 2052 AGCCAGGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2099
DB 1258 AGCCAGTGGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
OY 2100 AGCTTCCAGGCGCCGAGTGTAT-----CCGCTGTGGAGATCCAGCTG 2144
DB 1318 AGCATCCAGATCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377
OY 2145 GCCAGAGATGGAGCTTTTCTTTCTTTGCGCCGCTTCCAGAGTCCAAAGACACCTCCCT 2204
DB 1378 GCTTCACATGAAGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1435
OY 2205 CCAGGGTCTCTCTTCCAGAGTGGCGGCGCCACTCAGCGCCGAGACCAACCACTCAC 2264
DB 1436 CCAAGACTTCTCTTCCAGAGTGGCGGCGCCACTCCTTCAAGGAGCAATGGGCTCAC 1495
OY 2265 CTCTGACCCCAAGTAAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2324
DB 1496 TCCC---ACCCAGTAAATATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1552
OY 2325 TGATGGATGCTCTTAAATATTAAGATGTTTGAAT 2363
DB 1553 GCA---GATGCTCTTAAATATTAAGATGTTTGAAT 1588

```

```

RESULT 8
US-09-949-016-13053
; Sequence 13053, Application US/09949016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: PasteSeq for Windows Version 4.0
; SEQ ID NO 13053
; LENGTH: 30337
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(30337)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13053

```



```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 21784
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(21784)
; OTHER INFORMATION: n = A,T,C or G
US-09-820-002-3
```

```
Query Match      13.8%; Score 325; DB 4; Length 21784;
Best Local Similarity 100.0%; Pred. No. 4.7e-71;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2039 AGACTCATCCGAGACCGCCGATGATGACCCAGCTTGACCGGTCTTCTGCTCG 2098
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 19460 AGACTCATCCGAGACCGCCGATGATGACCCAGCTTGACCCAGGTCTTCTGCTCG 19519
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2099 CAGCTCCAGGGCCCGAGGTGATCCGGTGGTGGATCAGCGTGGCCGAGATGGAGC 2158
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 19520 CAGCTCCAGGGCCCGAGGTGATCCGGTGGTGGATCAGCGTGGCCGAGATGGAGC 19579
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2159 GTTTTCTCTTCTGCGCCCGCTCCACAGGTCCAGAGCACCTTCCCTCCAGGGTCTCT 2218
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 19580 GTTTTCTCTTCTGCGCCCGCTCCACAGGTCCAGAGCACCTTCCCTCCAGGGTCTCT 19639
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2219 TCACAGTGGCGGGCCCACTAGACGCCGAGACCACTCACTTCTTCTGACCCCAT 2278
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 19640 TCACAGTGGCGGGCCCACTAGACGCCGAGACCACTCACTTCTTCTGACCCCAT 19699
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2279 GTAATATATTGCTGTGCTGTGAGACTCTGTCTAGAGTCCCTGATGATGGATGCTCT 2338
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 19700 GTAATATATTGCTGTGCTGTGAGACTCTGTCTAGAGTCCCTGATGATGGATGCTCT 19759
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2339 TTAATATAATAAGATGTTTGATT 2363
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 19760 TTAATATAATAAGATGTTTGATT 19784
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 11
US-09-820-002-15
; Sequence 15, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gam, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: C1001194
; CURRENT APPLICATION NUMBER: US/09/820.002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 601
; TYPE: DNA
; ORGANISM: HUMAN
US-09-820-002-15
```

```
Query Match      8.1%; Score 192; DB 4; Length 601;
Best Local Similarity 90.3%; Pred. No. 2.5e-38;
Matches 204; Conservative 1; Mismatches 21; Indels 0; Gaps 0;
```

```
QY 1411 GACTGGGTCTGACAGCGCCGACCTTCCCGAGCGGAGACCGGGTCTGTCCCGATGG 1470
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 173 GCGAGGTGGCCACCTCCACCCCTTCCCTGTGATGGAGGAGACCGGGTCTGTCCCGATGG 232
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1471 CGAGTGTGGCGGTGCGGTGGCCAGGCTTCTCCCAAGTCTGACGCTGGGGGTGAG 1530
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 233 CGAGTGTGGCGGTGCGGTGGCCAGGCTTCTCCCAAGTCTGACGCTGGGGGTGAG 292
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1531 GCTGTGCTTACACAGCGGGGCTATCTTCCCTTGGGAGCCCAAGAGAGAGAACAGC 1590
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 293 GCTGTGCTTACACAGCGGGGCTATCTTCCCTTGGGAGCCCAAGAGAGAGAACAGC 352
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1591 AACGATATTGCCCTGTGCTCACTCTCCAGTCCCTGCCCCCTACAG 1636
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 353 AACGATATTGCCCTGTGCTCACTCTCCAGTCCCTGCCCCCTACAG 398
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 12
US-09-949-016-45925
; Sequence 45925, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949.016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45925
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-45925
```

```
Query Match      8.1%; Score 192; DB 4; Length 601;
Best Local Similarity 90.3%; Pred. No. 2.5e-38;
Matches 204; Conservative 1; Mismatches 21; Indels 0; Gaps 0;
```

```
QY 1411 GACTGGGTCTGACAGCGCCGACCTTCCCGAGCGGAGACCGGGTCTGTCCCGATGG 1470
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 173 GCGAGGTGGCCACCTCCACCCCTTCCCTGTGATGGAGGAGACCGGGTCTGTCCCGATGG 232
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1471 CGAGTGTGGCGGTGCGGTGGCCAGGCTTCCCGAGGCTGACGCTGGGGGTGAG 1530
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 233 CGAGTGTGGCGGTGCGGTGGCCAGGCTTCCCGAGGCTGACGCTGGGGGTGAG 292
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1531 GCTGTGCTTACACAGCGGGGCTATCTTCCCTTGGGAGCCCAAGAGAGAGAACAGC 1590
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 293 GCTGTGCTTACACAGCGGGGCTATCTTCCCTTGGGAGCCCAAGAGAGAGAACAGC 352
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1591 AACGATATTGCCCTGTGCTCACTCTCCAGTCCCTGCCCCCTACAG 1636
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 353 AACGATATTGCCCTGTGCTCACTCTCCAGTCCCTGCCCCCTACAG 398
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 13
US-09-820-002-16
; Sequence 16, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gam, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: C1001194
; CURRENT APPLICATION NUMBER: US/09/820.002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
```

```

; SOFTWARE: PaatSeq for Windows Version 4.0
; SEQ ID NO: 16
; LENGTH: 601
; TYPE: DNA
; ORGANISM: HUMAN
US-03-820-002-16

```

Query Match	8.1%;	Score 191;	DB 4;	Length 601;
Best Local Similarity	100.0%;	Pred. No. 4.4e-38;		
Matches 191; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1446 GCGGAACCGGGTCTCTGTCCCGATGCGAGGTTGGCCGGAGCCGAGGCGCTCTCC 1505

Db 16 GCGGAACCGGGTCTCTGTCCCGATGCGAGGTTGGCCGGAGCCGAGGCGCTCTCTCC 75

0y 150b CCACGGCTCAGCAGCTGGGGGTGCAGGGCTGAGTCTACACACGGGGCTATCTCCCTTGG 155b
 76 CCACGGCTCAGCAGCTGGGGGTGCAGGGCTGAGTCTACACACGGGGCTATCTCCCTTGG 135
 Db

0y 1566 GGACCCCAACAGCAGGAGAAACGACCAATATTGGCCCGGTCCACCTCTCCAGTCCCT 1625
 Db 136 GGACCCCAACAGCAGGAGAAACGACCAATATTGGCCCGGTCCACCTCTCCAGTCCCT 195

QY	1626	GCCCCTCACAG	1636
Db	196	GCCCCTCACAG	206

RESULT 14
US-09-820-002-14

```

: Patent No.6482630
: GENERAL INFORMATION:
: APPLICANT: Gan, Weiniu
: APPLICANT: Ye, Jane
: APPLICANT: Difrancesco, Valentina
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
: TITLE OF INVENTION: USES THEREOF
: FILE REFERENCE: C1001194
: CURRENT APPLICATION NUMBER: US/09/820,002 *
: CURRENT FILING DATE: 2001-03-29
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 14
: LENGTH: 601
: TYPE: DNA
: ORGANISM: HUMAN
: US-09-820-002-14

```

Query Match	7.2%	Score 169.4;	DB 4;	Length 601;
Best Local Similarity	-99.4%;	Pred. No. 1.2e-32;		
Matches 170; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	1278	AGACTGTGTGGCCGGCAGGAAGTCTCCCTGTGGACCCGATCTGTGGAGGGCCGGGACACCACTTT	1337
Db	320	AGACTGTGTGGCCGGCAGGAAGTCTCCCTGTGGACCCGATCTGTGGAGGGCCGGGACACCACTTT	379
QY	1338	GGGGCCGGTGGCCGGTGGCAATCAGCTTCGCTATGATGTGACCAACACTCTGTGGGGGATC	1397
Db	380	GGGGCCGGTGGCCGGTGGCAATCAGCTTCGCTATGATGTGACCAACACTCTGTGGGGGATC	439
QY	1398	CCTGCTCTCCGGGGACTGGGTGCTGACAGCCGCCACATGCTTCCCGGAGCG	1448
Db	440	CCTGCTCTCTCCGGGGACTGGGTGCTGACAGCCGCCACATGCTTCCCGGAGTG	490

```

RESULT 15
US-09-949-016-45924
; Sequence 45924, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

```

```

1  APPLICANT: VENTER, J. Craig et al.
2  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
3  TITLE OF INVENTION: WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF
4  FILE REFERENCE: CLO01307
5  CURRENT APPLICATION NUMBER: US/09/949,016
6  CURRENT FILING DATE: 2000-04-14
7  PRIOR APPLICATION NUMBER: 60/241,755
8  PRIOR FILING DATE: 2000-10-20
9  PRIOR APPLICATION NUMBER: 60/237,768
10 PRIOR FILING DATE: 2000-10-03
11 PRIOR APPLICATION NUMBER: 60/231,498
12 PRIOR FILING DATE: 2000-09-08
13 NUMBER OF SEQ ID NOS: 207012
14 SOFTWARE: FastSeq for Windows Version 4.0
15 SEQ ID NO 45924
16 LENGTH: 601
17 TYPE: DNA
18 ORGANISM: Human
19 US-09-949-016-45924

```

Query Match	7.2%	Score 169.4;	DB 4;	Length 601;
Best Local Similarity	99.4%;	Pred. No. 1.2e-32;		
Matches 170; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

Db	1278	133
Oy	AGACTGTGGCCCGCAGAGAACTGCTCCCGTGTGACCCGCATCGTGGGAGAGGCCGGGACACACAGCTT	133
324	AGACTGTGGCCCGCAGAGAACTGCTCCCGTGTGACCCGCATCGTGGGAGAGGCCGGGACACACAGCTT	383

Oy 1338 GGGCCCGGTGGCCGTGGCAAGTCAGACCTTCGTATGATGAGACACCTCTGTGGGGGATC 1339
 Db 384 GGGCCCGGTGGCCGTGGCAAGTCAGACCTTCGTATGATGAGACACACCTCTGTGGGGGATC 443

DY 1398 CCGTCTCCGGGAGCTGGTGCTGACACGCCCACTGCTTCGCCGAGCG 1448
| | | | | | | | | | | | | | | | | | | | | |
Db 444 CCGTCTCCGGGAGCTGGTGCTGACACGCCCACTGCTTCGCCGAGCG 494

Search completed: August 17, 2005, 10:50:38
Job time : 559 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2005, 01:11:50 ; Search time 1911 Seconds
(without alignments)
7319.910 Million cell updates/sec

Title: US-10-054-498-1

Perfect score: 2363
Sequence: 1 tcgagccgcgttcctccagsga.....taataagatcggtttgattc 2363

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2363	100.0	2363	6	ABL68569 Kidney ca
2	2363	100.0	2363	6	AAD40828 Human hep
3	2363	100.0	2363	6	AAD40647 Human hep
4	2363	100.0	2363	6	ABN97195 Gene #369
5	2363	100.0	2363	6	ABN97195 Gene #369
6	2363	100.0	2363	12	ADL56829 Human CDV
7	2363	100.0	2363	12	ADL56829 Human CDV
8	2303.4	97.5	2311	8	ACC46451 Human cdn
9	1638.6	69.3	1903	13	ACN41149 Human dit
10	1593.6	67.4	1783	4	AAD13167 Human hep
11	1593.6	67.4	1783	5	AAS43103 CDNA enco
12	1593.6	67.4	1783	6	ABO79532 Human hep
13	1593.6	67.4	1783	9	ADAI7045 Human enco
14	1593.6	67.4	1783	10	ADAI7045 CDNA enco
15	1593.6	67.4	1783	10	ADD18428 Prostate
16	1593.6	67.4	1783	10	ADD18428 Human pro
17	1593.6	67.4	1783	11	ADN39842 Human cel
18	1593.6	67.4	1783	12	ADJ46927 Cancer/an
19	1593.6	67.4	1783	13	ADRI8849 Human tra
20	1547.6	65.5	1811	13	ACN41153 Human dia

ALIGNMENTS

21	1493.4	63.2	1758	13	ACN41152	ACN41152 Human dia
22	1482	62.7	1482	6	AAS62259	AAS62259 CDNA sequ
23	1446	61.2	1706	10	ADJ57076	ADJ57076 Human pro
24	1405	59.5	1769	12	ADJ57605	ADJ57605 Human hep
25	1390	58.8	1763	13	ACN41151	ACN41151 Human dia
26	1349	57.1	1794	13	ACN41150	ACN41150 Human dia
27	1269	53.7	1615	10	AAI62370	AAI62370 Human pro
28	1269	53.7	1615	10	ABX16007	ABX16007 Human cdn
29	1254	53.1	1254	12	ADL59245	ADL59245 DNA enco
30	1098.6	46.5	1671	13	ACN41154	ACN41154 Human dia
31	1097.8	46.5	7033	12	ADJ07393	ADJ07393 Modified
32	1097	46.4	12242	12	ADJ07394	ADJ07394 Modified
33	1031	43.6	1165	12	ADJ07088	ADJ07088 Human pro
34	1029	43.5	1584	12	ADJ06886	ADJ06886 CDNA enco
35	1023.6	43.3	1739	10	ADJ59230	ADJ59230 Toxicity-
36	1023.6	43.3	1739	10	ADJ52781	ADJ52781 Primary r
37	1008.2	42.7	1605	2	AAI15134	AAI15134 CDNA enco
38	1008.2	42.7	1605	2	AAZ31883	AAZ31883 Mouse hep
39	858.4	36.3	2175	11	ADJ01874	ADJ01874 Human cdn
40	777.2	32.9	1192	12	ADJ07395	ADJ07395 Modified
41	449.4	19.0	644	13	ADJ03467	ADJ03467 Novel can
42	402.8	17.0	493	9	ACH45550	ACH45550 Human foe
43	391.2	16.6	614	6	ABX30301	ABX30301 Human G-P
44	380.2	16.1	497	10	ABX08968	ABX08968 CDNA enco
45	374.8	15.9	494	9	ACH45563	ACH45563 Human foe

RESULT 1
ABL68569
ID ABL68569 standard; DNA; 2363 BP.
XX AC ABL68569;
XX AC
DT 15-MAY-2002 (first entry)
XX DE Kidney cancer related gene sequence SEQ ID NO:6906.
XX KM Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KM stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;
KM cytosolic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
KM gene; ds.
XX OS Homo sapiens.
XX OS
XX PN WO200194629-A2.
XX PD 13-DEC-2001.
XX PF 30-MAY-2001; 2001WO-US010838.
XX PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.

20030165839

PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX (AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
PS Claim 1; SEQ ID NO 6906; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in AB161664
CC to AB170110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms
CC tumour
XX
SQ Sequence 2363 BP; 403 A; 808 C; 696 G; 456 T; 0 U; 0 Other;
Query Match 100.0%; Score 2363; DB 6; Length 2363;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCGAGCCGCTTTCCAGGAGCCCTAAGTGAAGGCGCAAGGTGAAGGAGCTGAGCTTAC 60
DB 1 TCGAGCCGCTTTCCAGGAGCCCTAAGTGAAGGCGCAAGGTGAAGGAGCTGAGCTTAC 60
QY 61 AGGCCCCAGCCAGCCGCTTGCCTCCAGGCGCGCCGCTGCTGGGGGCCACATGCTCC 120
DB 61 AGGCCCCAGCCAGCCGCTTGCCTCCAGGCGCGCCGCTGCTGGGGGCCACATGCTCC 120
QY 121 TGGCCAGGCTGGAAGCTGACCCGAGCCCGGCACTAAGTGAAGGCTCCGCCCCACATCG 180
DB 121 TGGCCAGGCTGGAAGCTGACCCGAGCCCGGCACTAAGTGAAGGCTCCGCCCCACATCG 180
QY 181 TGAACCCAGGTAAGGACAAAGGGCCCCAGACTACAGTTCAGCCCTGAGAGCAGGGG 240

DB 181 TGAACCCAGGTAAGGACAAAGGGCCCCAGACTACAGTTCAGCCCTGAGAGCAGGGG 240
QY 241 TTCCCTATCCCCCACCACAGCTTAATGCCACCTCTAATAGAGGGGTTCTGGGAGCC 300
DB 241 TTCCCTATCCCCCACCACAGCTTAATGCCACCTCTAATAGAGGGGTTCTGGGAGCC 300
QY 301 TGAAGAGGGGCACTATAGAGTCTCCCAAGCACTAGTGTCTGCTGCTCTTCTT 360
DB 301 TGAAGAGGGGCACTATAGAGTCTCCCAAGCACTAGTGTCTGCTGCTCTTCTT 360
QY 361 CAGACTAGCCGTTGAGACCCCAAGTCTTTCCTCCCAAGCCAGAGTTCCAGCCCTAG 420
DB 361 CAGACTAGCCGTTGAGACCCCAAGTCTTTCCTCCCAAGCCAGAGTTCCAGCCCTAG 420
QY 421 GCCCTCTCTCCCTCATATCTAGGAGTCTGGGCCCCCAAAATTCCTCTTCCCAAGACTTA 480
DB 421 GCCCTCTCTCCCTCATATCTAGGAGTCTGGGCCCCCAAAATTCCTCTTCCCAAGACTTA 480
QY 481 TGATTTAGGTTCTCAGCTGTCTCTCCCTCAAAACGGAGTCTCAGTCCCTGCTCAC 540
DB 481 TGATTTAGGTTCTCAGCTGTCTCTCCCTCAAAACGGAGTCTCAGTCCCTGCTCAC 540
QY 541 CAGGCTCAGGCAATGGGGTCCCATCTCTGCAAAATCCAGGCGTCCCGCTGCTGTCA 600
DB 541 CAGGCTCAGGCAATGGGGTCCCATCTCTGCAAAATCCAGGCGTCCCGCTGCTGTCA 600
QY 601 GACACTGACCCCAATCTTGAACCCAGCCCAATCGCTCCGTATCAAGGCGGCTCTGG 660
DB 601 GACACTGACCCCAATCTTGAACCCAGCCCAATCGCTCCGTATCAAGGCGGCTCTGG 660
QY 661 CCAAGGCCAGTCCCTAAGCCTGCTGATGAGACGCTGAGACTGGGGGCGCAGAGACT 720
DB 661 CCAAGGCCAGTCCCTAAGCCTGCTGATGAGACGCTGAGACTGGGGGCGCAGAGACT 720
QY 721 GGGCTGGGCTGGGCTCCCGAGGCGCTGCTCCCTGCTCATCTCTCAAGGTTCCACC 780
DB 721 GGGCTGGGCTGGGCTCCCGAGGCGCTGCTCCCTGCTCATCTCTCAAGGTTCCACC 780
QY 781 TGGCCCGAGAGTCAAGCAGGAAATCTTAAAGAGGAGCTGACATGGGGCGAAGAG 840
DB 781 TGGCCCGAGAGTCAAGCAGGAAATCTTAAAGAGGAGCTGACATGGGGCGAAGAG 840
QY 841 GGTGGCGAGTGTGCTGCTCTCCAGACCAAGGTGACGCTCTGCGGGGACC 900
DB 841 GGTGGCGAGTGTGCTGCTCTCCAGACCAAGGTGACGCTCTGCGGGGACC 900
QY 901 CTGCTACTTGTGACAGCATCGGGGCGGCAATCTGAGGCTGTTCTCTCAGG 960
DB 901 CTGCTACTTGTGACAGCATCGGGGCGGCAATCTGAGGCTGTTCTCTCAGG 960
QY 961 AGTGAACAGGAGCGCTGTACCCAGTGCAGGTGACGCTGCGGAGCGCTGATGATG 1020
DB 961 AGTGAACAGGAGCGCTGTACCCAGTGCAGGTGACGCTGCGGAGCGCTGATGATG 1020
QY 1021 TTTGACAAAGACGGAAGGAGCGTGGCGGCTGCTCTCTGCGGCTTCAAGCGCAGGTA 1080
DB 1021 TTTGACAAAGACGGAAGGAGCGTGGCGGCTGCTCTCTGCGGCTTCAAGCGCAGGTA 1080
QY 1081 GCCGAGTCAAGTGCAGAGAGATGGGCTTCTCAAGGCACTGACCCAGGCTGAGC 1140
DB 1081 GCCGAGTCAAGTGCAGAGAGATGGGCTTCTCAAGGCACTGACCCAGGCTGAGC 1140
QY 1141 GTGCGAAGCGGCGGCGCAATGCAAGTGGGCTTCTCTGTGTGTGACGAGGGAGGCTG 1200
DB 1141 GTGCGAAGCGGCGGCGCAATGCAAGTGGGCTTCTCTGTGTGTGACGAGGGAGGCTG 1200
QY 1201 CCCCAACACCAAGAGGCTGTGAGAGTCAATCTCGGTGTGATTTGCCAGAGCGCTTTC 1260
DB 1201 CCCCAACACCAAGAGGCTGTGAGAGTCAATCTCGGTGTGATTTGCCAGAGCGCTTTC 1260
QY 1261 TTGGCGGCAATCTGCAAGACTGTGGCGGAGGAAGCTGCGGTGACCGCATGTGGGA 1320

Db 1261 TTGGCCGCATCTGCGAAGACTGTGAGCCGAGAAAGCTGCCCGTGGAACCGCATCGTGGGA 1320
 QY 1321 GGGCCGGGACACCAAGCTTGGGCGGCTGGCCGCAAGTCAAGCTTGGCTATGATGAGACA 1380
 Db 1321 GGGCCGGGACACCAAGCTTGGGCGGCTGGCCGCAAGTCAAGCTTGGCTATGATGAGACA 1380
 QY 1381 CACCTCTGTGGGGGATCCCTGCTCCGGGGGACTGGGGGTGTGACAGACCGGCCCACTGGCTTC 1440
 Db 1381 CACCTCTGTGGGGGATCCCTGCTCCGGGGGACTGGGGGTGTGACAGACCGGCCCACTGGCTTC 1440
 QY 1441 CCGAGCGGAAACCGGGGCTCTGTCCCGATGAGCGAGTGTGGCCGATGCCGTGGCCAGGCC 1500
 Db 1441 CCGAGCGGAAACCGGGGCTCTGTCCCGATGAGCGAGTGTGGCCGATGCCGTGGCCAGGCC 1500
 QY 1501 TCTCCCAACGGTCTGCAAGTGGGGGTGCAAGCTGTGTCTACCAAGCGGGGCTATCTTCC 1560
 Db 1501 TCTCCCAACGGTCTGCAAGTGGGGGTGCAAGCTGTGTCTACCAAGCGGGGCTATCTTCC 1560
 QY 1561 TTTGGGGAGCCCAACAGGAGGAGAAAGCAACGATATTGGCCGTGTCCACTCTCCAGT 1620
 Db 1561 TTTGGGGAGCCCAACAGGAGGAGAAAGCAACGATATTGGCCGTGTCCACTCTCCAGT 1620
 QY 1621 CCCCTGCCCTCACAGAAATACATCCAGCTGTGTGCTCTCCAGCTGCGGCGAGCCCTG 1680
 Db 1621 CCCCTGCCCTCACAGAAATACATCCAGCTGTGTGCTCTCCAGCTGCGGCGAGCCCTG 1680
 QY 1681 GTGATGCGAAGATCTGTACCGTGAAGCGGCTGGGGGCAACAGCAGTACTATGGCCACAG 1740
 Db 1681 GTGATGCGAAGATCTGTACCGTGAAGCGGCTGGGGGCAACAGCAGTACTATGGCCACAG 1740
 QY 1741 GCGGGGGTACTCCAGAGAGGCTCGAGTCCCAATATCACCAATGATGTGCAATGGCCCT 1800
 Db 1741 GCGGGGGTACTCCAGAGAGGCTCGAGTCCCAATATCACCAATGATGTGCAATGGCCCT 1800
 QY 1801 GACTTCTATGAAACCAAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGGGTGGC 1860
 Db 1801 GACTTCTATGAAACCAAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGGGTGGC 1860
 QY 1861 ATTGATGCTGCGCAGGCGCAAGCGGCTGCTCTTGTGTGTGAGAGACAGATCTTCGG 1920
 Db 1861 ATTGATGCTGCGCAGGCGCAAGCGGCTGCTCTTGTGTGTGAGAGACAGATCTTCGG 1920
 QY 1921 ACGCCAGGTTGGGGGCTGTGTGCAATTGTGAGTTGGGGCACTGGCTGTGCCCTGGCCAG 1980
 Db 1921 ACGCCAGGTTGGGGGCTGTGTGCAATTGTGAGTTGGGGCACTGGCTGTGCCCTGGCCAG 1980
 QY 1981 AAGCCAGGCGTCTACACCAAGATCAAGTCTTCGGGAGTGGATCTTCCAGGCCATAAG 2040
 Db 1981 AAGCCAGGCGTCTACACCAAGATCAAGTCTTCGGGAGTGGATCTTCCAGGCCATAAG 2040
 QY 2041 ACTCACTCCGAAGCGAGCGGATGATGACCCAGCTCTGACCGGTGGCTTCTCGTGGCA 2100
 Db 2041 ACTCACTCCGAAGCGAGCGGATGATGACCCAGCTCTGACCGGTGGCTTCTCGTGGCA 2100
 QY 2101 GCGTCCAGAGGCGCCAGAGTATCCCGGTGTGTGAGATCCAGCTGGGCGAGATGGAGCT 2160
 Db 2101 GCGTCCAGAGGCGCCAGAGTATCCCGGTGTGTGAGATCCAGCTGGGCGAGATGGAGCT 2160
 QY 2161 TTTTCTTCTTGGGCGGGTCTCACAGGTCCAGAGTCAACCTCTCCAGAGGCTCTCTCTTC 2220
 Db 2161 TTTTCTTCTTGGGCGGGTCTCACAGGTCCAGAGTCAACCTCTCCAGAGGCTCTCTCTTC 2220
 QY 2221 CACAGTGGCGGGCCCACTCAGCCCCGAGAACCAACCACTCAGCTCCTGACCCCAATGT 2280
 Db 2221 CACAGTGGCGGGCCCACTCAGCCCCGAGAACCAACCACTCAGCTCCTGACCCCAATGT 2280
 QY 2281 AAAATATTGTCTGCTGTGGAGACTCCCTGTAGAGTGGCCCTGTGATGATGGATGCTCTTT 2340
 Db 2281 AAAATATTGTCTGCTGTGGAGACTCCCTGTAGAGTGGCCCTGTGATGATGGATGCTCTTT 2340
 QY 2341 AAAATATTAAAGATGTTTGAAT 2363
 Db 2341 AAAATATTAAAGATGTTTGAAT 2363

RESULT 2
 AAD40828
 ID AAD40828 standard; DNA; 2363 BP.
 XX
 AC AAD40828;
 XX
 DT 30-OCT-2002 (first entry)
 XX
 DE Human hepsin DNA.
 XX
 KW Human; hepsin; antisense compound; antisense therapy; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 826..2079
 FT /*tag= a
 FT /product= "Human hepsin protein"
 XX
 PN WO200250247-A2.
 XX
 PD 27-JUN-2002.
 XX
 PF 14-DEC-2001; 2001WO-US048341.
 XX
 PR 20-DEC-2000; 2000US-00742482.
 XX
 PA (ISIS-) ISIS PHARM INC.
 XX
 PI Cowbert LM;
 XX
 DR WPI; 2002-519882/55.
 XX
 DR P-PSDB; AAE24850.
 XX
 PT Novel antisense compound targeted to nucleic acids encoding human hepsin,
 PT useful for inhibiting the expression of hepsin in human cells or tissues,
 PT and for treating humans having a disease associated with human hepsin.
 XX
 PS Example 13; Page 89-92; 100pp; English.
 XX
 CC The invention relates to antisense compounds, compositions and methods
 CC for modulating the expression of hepsin. The compositions comprise
 CC antisense compounds, particularly antisense oligonucleotides, targeted
 CC to nucleic acids encoding hepsin. The antisense compound is useful for
 CC inhibiting the expression of hepsin in human cells or tissues. It is also
 CC useful for treating an animal having a disease or condition associated
 CC with hepsin, by inhibiting expression of hepsin. It is useful for
 CC diagnostics, therapeutics, prophylaxis and as research reagents and kits.
 CC It is also used in antisense therapy. The present sequence is human
 CC hepsin DNA
 CC
 CC
 SQ Sequence 2363 BP; 403 A; 808 C; 696 G; 456 T; 0 U; 0 Other;
 Query Match 100.0%; Score 2363; DB 6; Length 2363;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TCGAGCCCGCTTTCAGAGGACCTTACTGAGGCGCCACAGTGAAGGAGCTGGCTTACG 60
 Db 1 TCGAGCCCGCTTTCAGAGGACCTTACTGAGGCGCCACAGTGAAGGAGCTGGCTTACG 60
 QY 61 AGGCCCAAGCCAGCGCTGTGCTCCAGGCGCGCTGCTGGGGGCAACATGTCTCC 120
 Db 61 AGGCCCAAGCCAGCGCTGTGCTCCAGGCGCGCTGCTGGGGGCAACATGTCTCC 120
 QY 121 TGGCCAGGCTTGAAGACTGACCGGACCGGCACTACTTGAAGGCTTCCGCCCACTGC 180
 Db 121 TGGCCAGGCTTGAAGACTGACCGGACCGGCACTACTTGAAGGCTTCCGCCCACTGC 180
 QY 181 TGGAGCCCAAGGATGAAGCAAGGCGCCGAGCTCAGTTCAGAGCTGAGGAGCAGGG 240
 Db 181 TGGAGCCCAAGGATGAAGCAAGGCGCCGAGCTCAGTTCAGAGCTGAGGAGCAGGG 240

Db 181 TGACCCCAAGGATAGGACAAGAGGCCCCAGACTCACAGTTCCAGCCCTGAGACAGAGG 240
Qy 241 TTCCCTCATCCCCCACCACCTTAATATGCCACTCTCTTAATAGAGGGTTCTTGGGACC 300
Db 241 TTCCCTCATCCCCCACCACCTTAATATGCCACTCTCTTAATAGAGGGTTCTTGGGACC 300
Qy 301 TGAAGAGGGGSCATATGACGTCTCCCAAGACCTTAGTGTTCGTCTGCTCTTCCCTT 360
Db 301 TGAAGAGGGGSCATATGACGTCTCCCAAGACCTTAGTGTTCGTCTGCTCTTCCCTT 360
Qy 361 CAGACTCAGCGGTTGAGACCCCAAGTCCTTCTCCCAAGACCCAGAGTTCCAGCCCTCAG 420
Db 361 CAGACTCAGCGGTTGAGACCCCAAGTCCTTCTCCCAAGACCCAGAGTTCCAGCCCTCAG 420
Qy 421 GCGCCCTCCTCCTCATATCTAGGAGATCCTGGCCCCCAATTCCTCTTCCCAAGACTTA 480
Db 421 GCGCCCTCCTCCTCATATCTAGGAGATCCTGGCCCCCAATTCCTCTTCCCAAGACTTA 480
Qy 481 TGATTTGAGTCTCAGCTGTCTCTCCTCCCTCAAAACCGGGATCTCAGTCCCTGCTCAC 540
Db 481 TGATTTGAGTCTCAGCTGTCTCTCCTCCCTCAAAACCGGGATCTCAGTCCCTGCTCAC 540
Qy 541 CAGGCTCAGGATGAGGAGTCCCAATCCCTGCAAAATCCAGAGGTCCTCCCTGCTGTCA 600
Db 541 CAGGCTCAGGATGAGGAGTCCCAATCCCTGCAAAATCCAGAGGTCCTCCCTGCTGTCA 600
Qy 601 GACACTACCCCATCTTGAACCCAGCCCAATCTGCGTCCGTGATCAAGGAGTCTCTGG 660
Db 601 GACACTACCCCATCTTGAACCCAGCCCAATCTGCGTCCGTGATCAAGGAGTCTCTGG 660
Qy 661 CCAAGGCCCAGTCCCTACAGCTGCTGATGAGACGCTGAGGAGCTGGGGGCGCCAGACT 720
Db 661 CCAAGGCCCAGTCCCTACAGCTGCTGATGAGACGCTGAGGAGCTGGGGGCGCCAGACT 720
Qy 721 GGGGCTGGGCTGGGCTCCCCAGGCGCTGCTCCCTGCTCATCTCTCAAGGTCACCC 780
Db 721 GGGGCTGGGCTGGGCTCCCCAGGCGCTGCTCCCTGCTCATCTCTCAAGGTCACCC 780
Qy 781 TGGCCCAAGAGGTGAGCCAGGGAATCATTAACAAGGCACTGACATGAGGCGAAGAAG 840
Db 781 TGGCCCAAGAGGTGAGCCAGGGAATCATTAACAAGGCACTGACATGAGGCGAAGAAG 840
Qy 841 GGTGGCGGAGCTGTGACATGTGCTCCAGACCCAAAGTGGAGAGCTTCACTGCGGGACC 900
Db 841 GGTGGCGGAGCTGTGACATGTGCTCCAGACCCAAAGTGGAGAGCTTCACTGCGGGACC 900
Qy 901 CTGCTACTTCTGACAGCCATCGGGGCGGCAATCTGGGCAATTTGTGGCTTTCTCTCAGG 960
Db 901 CTGCTACTTCTGACAGCCATCGGGGCGGCAATCTGGGCAATTTGTGGCTTTCTCTCAGG 960
Qy 961 AGTACCAAGAGCGGCTGTATCCAGTGCAGGTCAAGTCTGAGACGCTCGGCTCATGGTC 1020
Db 961 AGTACCAAGAGCGGCTGTATCCAGTGCAGGTCAAGTCTGAGACGCTCGGCTCATGGTC 1020
Qy 1021 TTTTACAAGAGCGAGAGGAGCGTGGCGGCTGTGCTCTCGCCCTCCAAAGCCAGAGATA 1080
Db 1021 TTTTACAAGAGCGAGAGGAGCGTGGCGGCTGTGCTCTCGCCCTCCAAAGCCAGAGATA 1080
Qy 1081 GCCCGAGCTCAGCTGCGAGAGATGAGGCTTCTCAAGGGCACTGACCCCACTCGAAGCTGAC 1140
Db 1081 GCCCGAGCTCAGCTGCGAGAGATGAGGCTTCTCAAGGGCACTGACCCCACTCGAAGCTGAC 1140
Qy 1141 GTGGGAAGGAGGAGCGCAATGAGCAGTGGGGCTTCTTGTGTGTGTGTGTGTGTGTGTGT 1200
Db 1141 GTGGGAAGGAGGAGCGCAATGAGCAGTGGGGCTTCTTGTGTGTGTGTGTGTGTGTGTGT 1200
Qy 1201 CCCCCACCCAGAGAGCTGTGAGAGTCACTCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260
Db 1201 CCCCCACCCAGAGAGCTGTGAGAGTCACTCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260
Qy 1261 TTGGCCGCGCATCTGCGCAAGACTGTGTGGCGGAGAGGTCGCCGTGTGTGTGTGTGTGT 1320
Db 1261 TTGGCCGCGCATCTGCGCAAGACTGTGTGGCGGAGAGGTCGCCGTGTGTGTGTGTGTGT 1320

Qy 1321 GGGCGGGAACAACACTTTGGGCGGGTGGCGGTGGCAAGTCAAGCCTTGCTATGATGAGACA 1380
Db 1321 GGGCGGGAACAACACTTTGGGCGGGTGGCGGTGGCAAGTCAAGCCTTGCTATGATGAGACA 1380
Qy 1381 CACTCTGTGTGGGAGATCCCTGCTCTCCGGGAGTGTGGTGTGACAGCGGCCCACTGCTTC 1440
Db 1381 CACTCTGTGTGGGAGATCCCTGCTCTCCGGGAGTGTGGTGTGACAGCGGCCCACTGCTTC 1440
Qy 1441 CCGGAGCGGAACCGGGTCTGTCCCATGTGCGAGATGTTTTCGGGTGCGGCCCAAGGCC 1500
Db 1441 CCGGAGCGGAACCGGGTCTGTCCCATGTGCGAGATGTTTTCGGGTGCGGCCCAAGGCC 1500
Qy 1501 TCTCCCAAGGTCGACGCTGGGGGTGACAGTGTGCTTACCAAGGGAGGCTATCTTCCC 1560
Db 1501 TCTCCCAAGGTCGACGCTGGGGGTGACAGTGTGCTTACCAAGGGAGGCTATCTTCCC 1560
Qy 1561 TTTTGGGACCCCAACAGCGAGAGAACAGCAAGATATGCTCCGTGTCACTCTCCAGT 1620
Db 1561 TTTTGGGACCCCAACAGCGAGAGAACAGCAAGATATGCTCCGTGTCACTCTCCAGT 1620
Qy 1621 CCGCTGCGCTCACAAGATATCAGCTGTGTGCTCCAGCTGCGCGGCGCCGCTG 1680
Db 1621 CCGCTGCGCTCACAAGATATCAGCTGTGTGCTCCAGCTGCGCGGCGCCGCTG 1680
Qy 1681 GTGGATGGCAAGATCTGTACCGTGAAGGGCTGGGGGCAACGCGAGTACTATGGCCACAG 1740
Db 1681 GTGGATGGCAAGATCTGTACCGTGAAGGGCTGGGGGCAACGCGAGTACTATGGCCACAG 1740
Qy 1741 GCGGGGTACTCCAGAGAGGCTGAGTCCCATTAATACAGATATGTCTGCAATGGCGCT 1800
Db 1741 GCGGGGTACTCCAGAGAGGCTGAGTCCCATTAATACAGATATGTCTGCAATGGCGCT 1800
Qy 1801 GACTTCTATGGAACCAAGATCAAGCCCAATATGTTCTGTCTGTGCTTACCCCGAGGGTGC 1860
Db 1801 GACTTCTATGGAACCAAGATCAAGCCCAATATGTTCTGTCTGTGCTTACCCCGAGGGTGC 1860
Qy 1861 ATTGATGCTGCGCAGGCGCAGACGCGTGTGCTCCCTTGTGTGTGTGTGTGTGTGTGTGT 1920
Db 1861 ATTGATGCTGCGCAGGCGCAGACGCGTGTGCTCCCTTGTGTGTGTGTGTGTGTGTGTGT 1920
Qy 1921 ACGCCAGGTTGGCGGCTGTGTGSCATTTGTGAGGATGTGGGCACTGAGCTGTGGCCAG 1980
Db 1921 ACGCCAGGTTGGCGGCTGTGTGSCATTTGTGAGGATGTGGGCACTGAGCTGTGGCCAG 1980
Qy 1981 AAGCCAGGCGTCTACACCAAGTCAAGTCACTTCCGGGAGTGTGATCTTCCAGGCCATPAA 2040
Db 1981 AAGCCAGGCGTCTACACCAAGTCAAGTCACTTCCGGGAGTGTGATCTTCCAGGCCATPAA 2040
Qy 2041 ACTCACTCCGAAGCGAGCGCATGTGTGACCCAGCTGTGACCGGTGCTTCTGCTGTGGCA 2100
Db 2041 ACTCACTCCGAAGCGAGCGCATGTGTGACCCAGCTGTGACCGGTGCTTCTGCTGTGGCA 2100
Qy 2101 GCTTCAGAGGCGGAGGTATCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2160
Db 2101 GCTTCAGAGGCGGAGGTATCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2160
Qy 2161 TTTTCTTCTTGGGCGGGTCAAGATTCAGAGCAACCTTCTCCAGGGTCTCTTCTTTC 2220
Db 2161 TTTTCTTCTTGGGCGGGTCAAGATTCAGAGCAACCTTCTCCAGGGTCTCTTCTTTC 2220
Qy 2221 CACAGTGGCGGGCCACTCAGCCCGAGACCAACCAACTCAGCCCTCTGACCCCACTGT 2280
Db 2221 CACAGTGGCGGGCCACTCAGCCCGAGACCAACCAACTCAGCCCTCTGACCCCACTGT 2280
Qy 2281 AAATATTTGTTCTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2340
Db 2281 AAATATTTGTTCTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2340
Qy 2341 AAATATTAAGATGTGTTTGTATT 2363
Db 2341 AAATATTAAGATGTGTTTGTATT 2363

RESULT 3
AAD40647 standard; DNA; 2363 BP.
XX AAD40647;
XX 30-OCT-2002 (first entry)
XX Human hepsin DNA.
XX Human; hepsin; inflammation; tumour; gene therapy; cytostatic; gene; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 826..2079
XX /tag= a
XX /product= "Human hepsin"
XX WO200250248-A2.
XX 27-JUN-2002.
XX 14-DEC-2001; 2001WO-US048431.
XX 20-DEC-2000; 2000US-00742703.
XX (ISIS-) ISIS PHARM INC.
XX (ABBO) ABBOTT LAB.
XX Marcotte PA; Cowbert LM;
XX WPI; 2002-519883/55.
XX P-PSDB; AAE25075.
XX New antisense oligonucleotides that modulate (particularly inhibit) human
XX hepsin, useful for treating a disease or condition associated with the
XX expression of hepsin, e.g. inflammation or tumor growth.
XX Example 13; Page 88-91; 101pp; English.
XX The invention relates to an antisense compound 8-30 nucleobases in length
XX targeted to a nucleic acid molecule encoding human hepsin. The antisense
XX compound specifically hybridises with and inhibits the expression of
XX human hepsin. The antisense compound or the pharmaceutical composition is
XX useful for treating animals and humans having a disease or condition
XX associated with the expression of hepsin, e.g. inflammation or tumour
XX growth. The antisense compounds are useful also for diagnostics,
XX prophylaxis (e.g. to prevent or delay infection, inflammation or tumour
XX formation) or as research reagents and kits. The method is useful for
XX modulating, specifically inhibiting the expression of hepsin which may be
XX used in research, e.g. to distinguish between functions of various members
XX of a biological pathway. The invention is used in gene therapy. The
XX present sequence is human hepsin DNA
XX
XX Sequence 2363 BP; 403 A; 808 C; 696 G; 456 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 2363; DB 6; Length 2363;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

181 TGAAGCCGAGGTAAAGGACAAAGGCCCCCAGACTCAAGTTCCAGCCCTGAGAGCAGAGG 240
181 TGAAGCCGAGGTAAAGGACAAAGGCCCCCAGACTCAAGTTCCAGCCCTGAGAGCAGAGG 240
241 TTCCCTCATCCCCCAGCCAGCCAGCTTAATGCCACCTCTTAATAGAGGGTTCTTGAGAGC 300
241 TTCCCTCATCCCCCAGCCAGCCAGCTTAATGCCACCTCTTAATAGAGGGTTCTTGAGAGC 300
301 TGAAGAGGGGCACTATGAGGTCTTCCCAAGACCTTAGTGTCTGTCTCTTCTTCTT 360
301 TGAAGAGGGGCACTATGAGGTCTTCCCAAGACCTTAGTGTCTGTCTCTTCTTCTT 360
361 CAGACTGAGCCGTTTGAAGCCCAAGTCTTCTTCCCAAGACCTTAGTGTCTGTCTCTT 420
361 CAGACTGAGCCGTTTGAAGCCCAAGTCTTCTTCCCAAGACCTTAGTGTCTGTCTCTT 420
421 GCCCTCTCTCTCTCACTAGGAGTCTTGAGGAGTCTTGAGGAGTCTTGAGGAGTCTTG 480
421 GCCCTCTCTCTCTCACTAGGAGTCTTGAGGAGTCTTGAGGAGTCTTGAGGAGTCTTG 480
481 TGATTTAGGTCTCTAGCTGTCTCTCTCTCAACCGGAGTCTCTAGTCTCTCTCTCAC 540
481 TGATTTAGGTCTCTAGCTGTCTCTCTCTCAACCGGAGTCTCTAGTCTCTCTCTCAC 540
541 CAGGCTCAGGCTAGGGGGTCCCATCTCTGCAAAATCCAGGCTCTCTCTCTCTCTCTCA 600
541 CAGGCTCAGGCTAGGGGGTCCCATCTCTGCAAAATCCAGGCTCTCTCTCTCTCTCTCA 600
601 GACACTGACCCCATCTCTTGAACCGAGCCCAATCTGCTCTGATCAAGGCTCTCTG 660
601 GACACTGACCCCATCTCTTGAACCGAGCCCAATCTGCTCTGATCAAGGCTCTCTG 660
661 CCAAGGCCAGTCTCTCAAGCTCTCTGATGAGACGCTGAGACCTGAGGAGGCGCAGAGACT 720
661 CCAAGGCCAGTCTCTCAAGCTCTCTGATGAGACGCTGAGACCTGAGGAGGCGCAGAGACT 720
721 GGGCTGGGCTGGGCTCCCCAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
721 GGGCTGGGCTGGGCTCCCCAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
781 TGGCCGAGAGGTCTAGGAGGAGATCATTAACAAGAGGAGTGAATGAGGAGAGAG 840
781 TGGCCGAGAGGTCTAGGAGGAGATCATTAACAAGAGGAGTGAATGAGGAGAGAG 840
841 GGTGGCCGAGCTGTGCT 900
841 GGTGGCCGAGCTGTGCT 900
901 CTGCTACTTCTGACAGGATCGGGGCGGATCTCTGGGCTCTGCTCTCTCTCTCTCTCT 960
901 CTGCTACTTCTGACAGGATCGGGGCGGATCTCTGGGCTCTGCTCTCTCTCTCTCTCT 960
961 AGTGAACGAGAGCGCTGTATCCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1020
961 AGTGAACGAGAGCGCTGTATCCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1020
1021 TTTGACAAAGCGAGAGGAGTGTGGGCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
1021 TTTGACAAAGCGAGAGGAGTGTGGGCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
1081 GCCGAGTCTAGCTCTGAGAGAGATGGGCTTCTCTGAGGACATGACCTCTCTCTCTCT 1140
1081 GCCGAGTCTAGCTCTGAGAGAGATGGGCTTCTCTGAGGACATGACCTCTCTCTCTCT 1140
1141 GTGCGAAGCGGCGGCGCAATGAGAGTGTGGGCTTCTCTGAGGAGAGAGGCTG 1200
1141 GTGCGAAGCGGCGGCGCAATGAGAGTGTGGGCTTCTCTGAGGAGAGAGGCTG 1200
1201 CCCCACACCCAGAGGCTGTGAGATCATCTCTGATGATGATGATGATGATGATGATGAT 1260
1201 CCCCACACCCAGAGGCTGTGAGATCATCTCTGATGATGATGATGATGATGATGATGAT 1260

QY 1261 TTGGCCGCGCATCTGCGCAAGACTGTGCGCGAGAAAGTCCCGTGGACCGGATGCTGGGA 1320
DB 1261 TTGGCCGCGCATCTGCGCAAGACTGTGCGCGAGAAAGTCCCGTGGACCGGATGCTGGGA 1320
QY 1321 GGGCGGGAGACACAGACTTGGGCGGATGGCGGTGGCAAGTCAAGCTTTCGCTATGATGAGACA 1380
DB 1321 GGGCGGGAGACACAGACTTGGGCGGATGGCGGTGGCAAGTCAAGCTTTCGCTATGATGAGACA 1380
QY 1381 CACCTCTGTGGGGGATCCCTCTCTCCGGGAGCTGGGTGTGACAGAGCCCATCTGCTTC 1440
DB 1381 CACCTCTGTGGGGGATCCCTCTCTCCGGGAGCTGGGTGTGACAGAGCCCATCTGCTTC 1440
QY 1441 CCGGAGCGGAACCGGGTCTGTGCCGATGCGAGTGTGGCGGATGCCGTGGCCGACGACC 1500
DB 1441 CCGGAGCGGAACCGGGTCTGTGCCGATGCGAGTGTGGCGGATGCCGTGGCCGACGACC 1500
QY 1501 TCTCCCAACGCTGTGACAGCTGGGGGTGACAGCTGTGGTCTACACGCGGGGCTATCTTCC 1560
DB 1501 TCTCCCAACGCTGTGACAGCTGGGGGTGACAGCTGTGGTCTACACGCGGGGCTATCTTCC 1560
QY 1561 TTTTGGGAGCCCAACAGGAGAGAGAACAGACGATATGTGCCCTGTCTCACTCTCCAGT 1620
DB 1561 TTTTGGGAGCCCAACAGGAGAGAGAACAGACGATATGTGCCCTGTCTCACTCTCCAGT 1620
QY 1621 CCCCTGCCCTCAGAGAAATATCATCATCAGCTGTGTGCTCCAGCTGCGCGGACAGCCCTG 1680
DB 1621 CCCCTGCCCTCAGAGAAATATCATCATCAGCTGTGTGCTCCAGCTGCGCGGACAGCCCTG 1680
QY 1681 GTGGATGGCAAGATCTGTACCGGTGACGCGCTGGGGCAACAGCAGTACTATGAGCCAAAG 1740
DB 1681 GTGGATGGCAAGATCTGTACCGGTGACGCGCTGGGGCAACAGCAGTACTATGAGCCAAAG 1740
QY 1741 GCGCGGGTACTCCAGAGAGGCTCGAGTCCCAATATCAGCATGATGTCTGAAATGGGCT 1800
DB 1741 GCGCGGGTACTCCAGAGAGGCTCGAGTCCCAATATCAGCATGATGTCTGAAATGGGCT 1800
QY 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGTGCTACCCGAGGGTGGC 1860
DB 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGTGCTACCCGAGGGTGGC 1860
QY 1861 ATTGATGCTTGGCCAGGGGAGCAGCGGTGTCTCTTGTGTGAGAGACAGCATCTCTCGG 1920
DB 1861 ATTGATGCTTGGCCAGGGGAGCAGCGGTGTCTCTTGTGTGAGAGACAGCATCTCTCGG 1920
QY 1921 ACGCCAGCTTGGCGGCTGTGTGAGATTTGGAGTTGGGGCAGCTGTGTGCTTGGCCAG 1980
DB 1921 ACGCCAGCTTGGCGGCTGTGTGAGATTTGGAGTTGGGGCAGCTGTGTGCTTGGCCAG 1980
QY 1981 AAGCCAGGCGTCTACACCAAGTCAAGTACTTCGGGAGTGGATCTTCCAGGCCATTAAG 2040
DB 1981 AAGCCAGGCGTCTACACCAAGTCAAGTACTTCGGGAGTGGATCTTCCAGGCCATTAAG 2040
QY 2041 ACTGACTGCCAAGCAGCGGAGTGTGACCCAGCTCTGACCGGTGTCTTCTGTGCGCA 2100
DB 2041 ACTGACTGCCAAGCAGCGGAGTGTGACCCAGCTCTGACCGGTGTCTTCTGTGCGCA 2100
QY 2101 GCCCTCAGGGGCCAGAGTATCCCGGTGTGGATCCAGCTGTGGCCGAGATGGAGCT 2160
DB 2101 GCCCTCAGGGGCCAGAGTATCCCGGTGTGGATCCAGCTGTGGCCGAGATGGAGCT 2160
QY 2161 TTTTCTTCTTGGGCGCGGTCTCCAGAGTTCAGAGACACCTCTCTCAAGGTCTCTCTTC 2220
DB 2161 TTTTCTTCTTGGGCGCGGTCTCCAGAGTTCAGAGACACCTCTCTCAAGGTCTCTCTTC 2220
QY 2221 CACAGTGGCGGCGCCACTCAGCCCCGAGACCAACCTCAGCTCTCTCTGACCCCATGT 2280
DB 2221 CACAGTGGCGGCGCCACTCAGCCCCGAGACCAACCTCAGCTCTCTCTGACCCCATGT 2280
QY 2281 AAAATATTGTTCTGTCTGTGGAGCTCTCTGTCTAAGTGTGCTTGAATGATGATGCTCTTT 2340
DB 2281 AAAATATTGTTCTGTCTGTGGAGCTCTCTGTCTAAGTGTGCTTGAATGATGATGCTCTTT 2340
QY 2341 AAATATATAAGATGTTTGAAT 2363

DB 2341 AAATATATAAGATGTTTGAAT 2363
RESULT 4
ID ABRN97195 standard; DNA; 2363 BP.
AC ABRN97195;
DT 13-AUG-2002 (first entry)
XX
DE Gene #3693 used to diagnose liver cancer.
KM Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KM metastatic liver tumour; cytostatic; expression profile; disease state;
KM disease progression; drug toxicity; drug efficacy; drug metabolism.
OS Homo sapiens.
XX
PM W0200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001MO-US030589.
XX
PR 02-OCT-2000; 2000US-0237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer, hepatocellular
PT carcinoma or metastatic liver tumor in a patient, involves detecting the
PT level of expression of two or more genes in a liver tissue sample.
XX
PS Claim 1; SEQ ID NO 3693; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABRN93503-ABRN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cyostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2363 BP; 403 A; 808 C; 696 G; 456 T; 0 U; 0 Other;
Query Match 100.0%; Score 2363; DB 6; Length 2363;
Best local Similarity 100.0%; Pred. No. 0;
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCGAGCCCGCTTTCAGAGGACCTTACTGAGGGGCCACAGTGAAGGAGCTGTGCTAGC 60
DB 1 TCGAGCCCGCTTTCAGAGGACCTTACTGAGGGGCCACAGTGAAGGAGCTGTGCTAGC 60
QY 61 AGGCCCCACGACCGGCTGTGCTTCCAGGCGCGGCTGTGCGGGCCACATGCTCC 120
DB 61 AGGCCCCACGACCGGCTGTGCTTCCAGGCGCGGCTGTGCGGGCCACATGCTCC 120
QY 121 TGCCGAGGCTTGAAGACTGACCCGAGCCCGGACATCTGAGGCTTCCGCCCACTGCG 180
DB 121 TGCCGAGGCTTGAAGACTGACCCGAGCCCGGACATCTGAGGCTTCCGCCCACTGCG 180

QY 181 TGGACCCAGGGTAAAGGACCAAGGGGCCAGACTCAGATTCCAGCCCTGAGAGAGAGGG 240
Db 181 TGGACCCAGGGTAAAGGACCAAGGGGCCAGACTCAGATTCCAGCCCTGAGAGAGAGGG 240
QY 241 TTCCCTCATCCCCCAAGCCAGCTTAATGAGGGGTTCTTGAGAGC 300
Db 241 TTCCCTCATCCCCCAAGCCAGCTTAATGAGGGGTTCTTGAGAGC 300
QY 301 TGAAGAGGGGACATATGACGTCTCCCAAGACCTAGAGTGTCTGTCTCTCTT 360
Db 301 TGAAGAGGGGACATATGACGTCTCCCAAGACCTAGAGTGTCTGTCTCTCTT 360
QY 361 CAGACTAGCCGTTGAGACCCAGCTTCTCCCAAGACCCAGAGATTTCCAGCCCTAG 420
Db 361 CAGACTAGCCGTTGAGACCCAGCTTCTCCCAAGACCCAGAGATTTCCAGCCCTAG 420
QY 421 GCCCCTCTCTCATCTAGAGGAGTCTGAGCCCAATTCCTCTTCCCAAGACTTA 480
Db 421 GCCCCTCTCTCATCTAGAGGAGTCTGAGCCCAATTCCTCTTCCCAAGACTTA 480
QY 481 TGAATTCAGGTCCTCAGCTGTCTCTCCCTCAAAACCGGAGATCTCAGTCCCTGTCCAC 540
Db 481 TGAATTCAGGTCCTCAGCTGTCTCTCCCTCAAAACCGGAGATCTCAGTCCCTGTCCAC 540
QY 541 CAGGCTCAGGATAGGGGGTCCCATCTCCGAAATTCAGAGGTCCCCCGCTGTGTCA 600
Db 541 CAGGCTCAGGATAGGGGGTCCCATCTCCGAAATTCAGAGGTCCCCCGCTGTGTCA 600
QY 601 GACACTAGCCCATCTCTTGAACCCAGCCCAATTCGCTGATCAAGGAGTGTCTTGG 660
Db 601 GACACTAGCCCATCTCTTGAACCCAGCCCAATTCGCTGATCAAGGAGTGTCTTGG 660
QY 661 CCAAGGCCCACTCCCTACAGCTGTGCTGATGAGAGCTGAGGAGGAGCCAGAGACT 720
Db 661 CCAAGGCCCACTCCCTACAGCTGTGCTGATGAGAGCTGAGGAGGAGCCAGAGACT 720
QY 721 GGGCTGGGCTGGGCTCCCCAGGCGCTTCCCTGATCTCTCAAGAGTCCCAAGC 780
Db 721 GGGCTGGGCTGGGCTCCCCAGGCGCTTCCCTGATCTCTCAAGAGTCCCAAGC 780
QY 781 TGGCCAGAGAGGTAGCCAGGAGATCATTAACAAGAGGAGTGAATGGCCAGAGAGAG 840
Db 781 TGGCCAGAGAGGTAGCCAGGAGATCATTAACAAGAGGAGTGAATGGCCAGAGAGAG 840
QY 841 GGTGGCCGAGTGTGCTGTCTCCAGACCCAGAGTGTGAGACTCTCACTGCGGAGC 900
Db 841 GGTGGCCGAGTGTGCTGTCTCCAGACCCAGAGTGTGAGACTCTCACTGCGGAGC 900
QY 901 CTGCTACTTCTGACAGCCATGAGGGGAGCTTGTGGGCAATTTGTGTCTCTCAAG 960
Db 901 CTGCTACTTCTGACAGCCATGAGGGGAGCTTGTGGGCAATTTGTGTCTCTCAAG 960
QY 961 AGTACCAAGAGCGCTGTATCCAGTGAAGTGAAGTGTGAGAGCGCTCGGCTCATGTGTC 1020
Db 961 AGTACCAAGAGCGCTGTATCCAGTGAAGTGAAGTGTGAGAGCGCTCGGCTCATGTGTC 1020
QY 1021 TTTTGAACAAGAGGAGCGTGTGCGGCTGTGCTGTCTCTGCGCTCCAAAGCAGAGGTA 1080
Db 1021 TTTTGAACAAGAGGAGCGTGTGCGGCTGTGCTGTCTCTGCGCTCCAAAGCAGAGGTA 1080
QY 1081 GCCGAGCTCAGCTGTGAGAGAGATGAGGCTTCTCAGGAGCACTGACCTCGAGAGTGAGC 1140
Db 1081 GCCGAGCTCAGCTGTGAGAGAGATGAGGCTTCTCAGGAGCACTGACCTCGAGAGTGAGC 1140
QY 1141 GTGCGAAGCGGAGGAGCAATGAGCACTCGGAGCTTCTGTGTGTGAGAGAGAGAGGCTG 1200
Db 1141 GTGCGAAGCGGAGGAGCAATGAGCACTCGGAGCTTCTGTGTGTGAGAGAGAGAGGCTG 1200
QY 1201 CCCCACACCAAGAGCTGTGAGAGTCACTCTCGTGTGTGATTTGCCAGAGGCGCTTTC 1260
Db 1201 CCCCACACCAAGAGCTGTGAGAGTCACTCTCGTGTGTGATTTGCCAGAGGCGCTTTC 1260

QY 1261 TTGGCCGCACTGTGCAAGACTGTGTGGCCGAGAGGCTGCCGTGTGAGCCGAGCTGTGGA 1320
Db 1261 TTGGCCGCACTGTGCAAGACTGTGTGGCCGAGAGGCTGCCGTGTGAGCCGAGCTGTGGA 1320
QY 1321 GGGCGGAGCACAGCTTGGGCGGATGTGGCGTGGCAAGTCAAGCTTGTGATATGAGAGCA 1380
Db 1321 GGGCGGAGCACAGCTTGGGCGGATGTGGCGTGGCAAGTCAAGCTTGTGATATGAGAGCA 1380
QY 1381 CACCTGTGTGGGAGTCCCTGTCTCCGAGGACCTGTGTGACAGCCGCCCATCTGTTTC 1440
Db 1381 CACCTGTGTGGGAGTCCCTGTCTCCGAGGACCTGTGTGACAGCCGCCCATCTGTTTC 1440
QY 1441 CCGAGCGGAACCGGAGTCTCCGATGAGGAGGTTTGGCGGAGCGGTGGCCCAAGGCC 1500
Db 1441 CCGAGCGGAACCGGAGTCTCCGATGAGGAGGTTTGGCGGAGCGGTGGCCCAAGGCC 1500
QY 1501 TCTCCCAAGCTGTGACGTGGGAGTGAAGCTGTGTCTCAACAGGAGGCTATTTTCC 1560
Db 1501 TCTCCCAAGCTGTGACGTGGGAGTGAAGCTGTGTCTCAACAGGAGGCTATTTTCC 1560
QY 1561 TTTGAGGACCCCAACAGGAGAGAACAGACATATTTGCTGTCTCACTCTCAAGT 1620
Db 1561 TTTGAGGACCCCAACAGGAGAGAACAGACATATTTGCTGTCTCACTCTCAAGT 1620
QY 1621 CCCCCTGCCCCCAAGAAATATCATCAGCTGTGTGCTCCAGCTGCCGAGGAGCCCTG 1680
Db 1621 CCCCCTGCCCCCAAGAAATATCATCAGCTGTGTGCTCCAGCTGCCGAGGAGCCCTG 1680
QY 1681 GTGATGAGCAAGATCTGTACCGGTAGCGGCTGTGGGCAACAGCACTATATGAGCCACAG 1740
Db 1681 GTGATGAGCAAGATCTGTACCGGTAGCGGCTGTGGGCAACAGCACTATATGAGCCACAG 1740
QY 1741 GCCGAGGACTCTCAGAGAGCTGAGTCCCAATTAACAGATATGTCTGCAATGAGCCT 1800
Db 1741 GCCGAGGACTCTCAGAGAGCTGAGTCCCAATTAACAGATATGTCTGCAATGAGCCT 1800
QY 1801 GACTTCTATGAGAAACAGATCAAGCCAGATGTTCTGTGTGCTACCCGAGGAGTGGC 1860
Db 1801 GACTTCTATGAGAAACAGATCAAGCCAGATGTTCTGTGTGCTACCCGAGGAGTGGC 1860
QY 1861 ATTGATGCTGTGCAAGGAGCTGAGTCCCTTGTGTGTGAGAGCAGCATCTTGG 1920
Db 1861 ATTGATGCTGTGCAAGGAGCTGAGTCCCTTGTGTGTGAGAGCAGCATCTTGG 1920
QY 1921 ACGCCAGTGTGGCGGCTGTGTGAGATGTGAGTGGGACCTGTGCTGCTGAGCCAG 1980
Db 1921 ACGCCAGTGTGGCGGCTGTGTGAGATGTGAGTGGGACCTGTGCTGCTGAGCCAG 1980
QY 1981 AAGCCAGGCTGTCAACCAAGTGAATCTTCCGAGAGTGGATCTTCCAGGCCATTAAG 2040
Db 1981 AAGCCAGGCTGTCAACCAAGTGAATCTTCCGAGAGTGGATCTTCCAGGCCATTAAG 2040
QY 2041 ACTCACTCCGAAGCGAGGAGTGAATCCAGCTGTGACCGGAGGCTTCTGCTGAGCA 2100
Db 2041 ACTCACTCCGAAGCGAGGAGTGAATCCAGCTGTGACCGGAGGCTTCTGCTGAGCA 2100
QY 2101 GCTCTCAAGGCGCAGAGTGAATCCGAGTGTGTGAGATCAACGCTGTGGCGAGAGATGAGAGT 2160
Db 2101 GCTCTCAAGGCGCAGAGTGAATCCGAGTGTGTGAGATCAACGCTGTGGCGAGAGATGAGAGT 2160
QY 2161 TTTTCTTCTTGTGGGCGGCTCAACAGGTTCMAAGACACCTTCTCTCAGGGTCTCTTTC 2220
Db 2161 TTTTCTTCTTGTGGGCGGCTCAACAGGTTCMAAGACACCTTCTCTCAGGGTCTCTTTC 2220
QY 2221 CACAGTGGCGGAGCCCATCAAGCCCGAGAACCAACCACTCACTCTGACCCCAATGT 2280
Db 2221 CACAGTGGCGGAGCCCATCAAGCCCGAGAACCAACCACTCACTCTGACCCCAATGT 2280
QY 2281 AAATATTTGTCTGTGTCTGTGAGTCTCTGTCTAGAGTCCCTGTATGATGAGATCTCTT 2340
Db 2281 AAATATTTGTCTGTGTCTGTGAGTCTCTGTCTAGAGTCCCTGTATGATGAGATCTCTT 2340
QY 2341 AAATTAATAAGATGTTTGAAT 2363

DB 1643 GGGCTGGGCTGGGCTCCCCAGGCGCTGCTCCCTCCATCTTCTTCAAGGTTCCACCC 1584
QY 791 TGGCCAGAGAGTCAAGGGAATCATTAACAGAGGAGTGAATGAGCGCAGAGAG 840
DB 1583 TGGCCAGAGAGTCAAGGGAATCATTAACAGAGGAGTGAATGAGCGCAGAGAG 1524
QY 841 GGTGGCCGAGCTGTGCTCAATGCTGCTCAAGCCCAAGGTGGAGCTCTCACTGGCGGAGC 900
DB 1523 GGTGGCCGAGCTGTGCTCAATGCTGCTCAAGCCCAAGGTGGAGCTCTCACTGGCGGAGC 1464
QY 901 CTGCTACTTCTGACAGCCATCGGGGCGGAGATCTGGGCGAATTGGAGCTTCTCTCAGG 960
DB 1463 CTGCTACTTCTGACAGCCATCGGGGCGGAGATCTGGGCGAATTGGAGCTTCTCTCAGG 1404
QY 961 AGTACCAAGAGCCGCTGTATCCAGTGAAGTCAAGCTGTGGAGAGCTCGAGTCAATGATC 1020
DB 1403 AGTACCAAGAGCCGCTGTATCCAGTGAAGTCAAGCTGTGGAGAGCTCGAGTCAATGATC 1344
QY 1021 TTGACCAAGAGGGAAGGAGCTGTGGGCTGTGTCTCTCGCGCTCAAGCGCCAGGTA 1080
DB 1343 TTGACCAAGAGGGAAGGAGCTGTGGGCTGTGTCTCTCGCGCTCAAGCGCCAGGTA 1284
QY 1081 GCGGAGCTCAGCTGAGAGAGATGGGCTTCCCTCAGGGGAGTGAACCACTCGAGCTGAGC 1140
DB 1283 GCGGAGCTCAGCTGAGAGAGATGGGCTTCCCTCAGGGGAGTGAACCACTCGAGCTGAGC 1224
QY 1141 GTGCGAAGCGGCGGCGCCATGAGCACTGAGGCTTCTTCTGTGTGAAGAGAGGAGCTG 1200
DB 1223 GTGCGAAGCGGCGGCGCCATGAGCACTGAGGCTTCTTCTGTGTGAAGAGAGGAGCTG 1164
QY 1201 CCCCACACCAAGAGCTGTGAGAGTCACTTCCGTGTGATTTGCCAGAGGCGGTTTC 1260
DB 1163 CCCCACACCAAGAGCTGTGAGAGTCACTTCCGTGTGATTTGCCAGAGGCGGTTTC 1104
QY 1261 TTGGCCGCGCATCTGCGAAGCTGTGGCGGAGAGAGTCCCGTGGAGCCGATGTTGGGA 1320
DB 1103 TTGGCCGCGCATCTGCGAAGCTGTGGCGGAGAGAGTCCCGTGGAGCCGATGTTGGGA 1044
QY 1321 GCGCGGAGCACAGCTTGGGCGGAGTGGCGGTGGCAAGTCAAGCTTCCGTATGATGAGCA 1380
DB 1043 GCGCGGAGCACAGCTTGGGCGGAGTGGCGGTGGCAAGTCAAGCTTCCGTATGATGAGCA 984
QY 1381 CACCTGTGTGGGAGTCCCTGCTCTCCGCGGAGCTGGGTGTGACAGCCGCCCATCTGCTTC 1440
DB 983 CACCTGTGTGGGAGTCCCTGCTCTCCGCGGAGCTGGGTGTGACAGCCGCCCATCTGCTTC 924
QY 1441 CCGGAGCGGAGCACCGGCTCTGCTCCCGATGGAGAGTGTGGCCGAGCCGTGGCCCAAGCC 1500
DB 923 CCGGAGCGGAGCACCGGCTCTGCTCCCGATGGAGAGTGTGGCCGAGCCGTGGCCCAAGCC 864
QY 1501 TCTCCCAAGGCTGTGACAGTGGGAGTGAAGCTGTGCTCAACAGGAGGCTATCTTCC 1560
DB 863 TCTCCCAAGGCTGTGACAGTGGGAGTGAAGCTGTGCTCAACAGGAGGCTATCTTCC 804
QY 1561 TTTTGGAGACCCCAACAGGAGAGAGCAAGCAATATTCCTGTGCTCACTCTCCAGT 1620
DB 803 TTTTGGAGACCCCAACAGGAGAGAGCAAGCAATATTCCTGTGCTCACTCTCCAGT 744
QY 1621 CCGCTGCCCTCAAGAAATATCAAGCTGTGGCTCCCAAGTGGCGGCGGAGGCTG 1680
DB 743 CCGCTGCCCTCAAGAAATATCAAGCTGTGGCTCCCAAGTGGCGGCGGAGGCTG 684
QY 1681 GTGATGGCAAGATCTGTACGTGAGCGGCTGGGCAACAGAGTATATGAGCAAG 1740
DB 683 GTGATGGCAAGATCTGTACGTGAGCGGCTGGGCAACAGAGTATATGAGCAAG 624
QY 1741 GCGGAGGATCTTCAGAGGCTTCGAGTCCCATATACCAATGATGTCTGCAATGGCGCT 1800
DB 623 GCGGAGGATCTTCAGAGGCTTCGAGTCCCATATACCAATGATGTCTGCAATGGCGCT 564
QY 1801 GACTTCTATGGAACCAAGATCAAGCCCAAGATGTTCTGTGTGCTACCCCGAGGCTGCG 1860

DB 563 GACTTCTATGGAACCAAGATCAAGCCCAAGATGTTCTGTGTGCTGACTACCCGAGGCTGCG 504
QY 1861 ATTATGCTGTCCAGGCGGAGCAGCGGTGTGCTCTTTGTGTGTGAGACAGATCTCTCGG 1920
DB 503 ATTATGCTGTCCAGGCGGAGCAGCGGTGTGCTCTTTGTGTGTGAGACAGATCTCTCGG 444
QY 1921 ACGCAGCTGGGCGGCTGTGTGAGATTTGAGTTGGGCACTATGCTGTGCTGCTGCGCCAG 1980
DB 443 ACGCAGCTGGGCGGCTGTGTGAGATTTGAGTTGGGCACTATGCTGTGCTGCTGCGCCAG 384
QY 1981 AAGCCAGGCGCTCAACCAAGATGAGTCTTCCGAGAGTGAATCTTCCAGGCCATAAAG 2040
DB 383 AAGCCAGGCGCTCAACCAAGATGAGTCTTCCGAGAGTGAATCTTCCAGGCCATAAAG 324
QY 2041 ACTCACTCCGAGGCGGAGGATGAGTCCAGCTTGAACCGGTGCTTCTGCTGCGCA 2100
DB 323 ACTCACTCCGAGGCGGAGGATGAGTCCAGCTTGAACCGGTGCTTCTGCTGCGCA 264
QY 2101 GCGTCCAGGCGGCGGAGGATCCGCTGTGTGGATTCACCGCTGGGCGGAGATGGAGCT 2160
DB 263 GCGTCCAGGCGGCGGAGGATCCGCTGTGTGGATTCACCGCTGGGCGGAGATGGAGCT 204
QY 2161 TTTTCTTCTTGGGCGGCTCCAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2220
DB 203 TTTTCTTCTTGGGCGGCTCCAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 144
QY 2221 CACAGTGGCGGCGGCGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2280
DB 143 CACAGTGGCGGCGGCGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 84
QY 2281 AAATATTTGTTGTGCTGTGCTGAGGACTCTGTGTAGTGTGCTGTGTAGTGTGCTGTG 2340
DB 83 AAATATTTGTTGTGCTGTGCTGAGGACTCTGTGTAGTGTGCTGTGTAGTGTGCTGTG 24
QY 2341 AAATATTAAGATGTTTGTGAT 2363
DB 23 AAATATTAAGATGTTTGTGAT 1
DB 23 AAATATTAAGATGTTTGTGAT 1
RESULT 6
ADL56829/c
ID ADL56829 standard; cDNA; 2363 BP.
XX ADL56829;
AC
XX
DT 20-MAY-2004 (first entry)
XX
DE Human NOVX cDNA #8.
XX
KW Human; NOVX; gene; 88; angiogenic-associated disorder; cancer;
cardiovascular disease; psoriasis; wound healing; stroke; cardiomyopathy;
atherosclerosis; cell signal processing disorder;
metabolic pathway modulation disorder; diabetes;
hyperproliferative disease; cirrhosis; keloid; psoriasis; osteoarthritis;
haemorrhage; ischaemic heart disease; renal disease; thrombosis;
hypertension; hypothyroidism; autoimmune disorder; multiple sclerosis;
systemic lupus erythematosus; rheumatoid arthritis;
autoimmune pulmonary inflammation; Guillain-Barre syndrome;
autoimmune thyroiditis; myasthenia gravis; graft-versus-host disease;
asthma; periodontal disease; Alzheimer's disease; Parkinson's disease;
Huntington's disease; amyotrophic lateral sclerosis; Shy-drager syndrome.
OS Homo sapiens.
XX
PN US2003215449-A1.
XX
PD 20-NOV-2003.
XX
PF 15-MAR-2002; 2002US-00099322.
XX
PR 11-JAN-2001; 2001US-0261013P.
PR 11-JAN-2001; 2001US-0261014P.
PR 11-JAN-2001; 2001US-0261018P.

PR 11-JAN-2001; 2001US-0261026P.
PR 11-JAN-2001; 2001US-0261029P.
PR 23-MAR-2001; 2001US-0278152P.
PR 17-AUG-2001; 2001US-0313170P.
PR 10-SEP-2001; 2001US-0318410P.
PR 11-JAN-2002; 2002US-00044554.

XX
PA (MEZE/) MEZES P D.
PA (RAST/) RASTELLI L.
PA (HERR/) HERRMANN J L.
PA (MACD/) MACDOUGALL J R.
PA (ZHON/) ZHONG H.
PA (CASMAN/) CASMAN S J.
PA (BOLD/) BOLDOG F L.
PA (SHIM/) SHIMKETS R A.
PA (GORM/) GORMAN L.
PA (EISEN/) EISEN A J.
PA (SPAD/) SPADERNA S K.
PA (VERN/) VERNET C A M.
PA (BERG/) BERGHS C.
PA (SPYT/) SPYTEK K A.
PA (DIP/) DIPITPO V A.
PA (ZERN/) ZERNHUSEN B D.
PA (PEYM/) PEYMAN J A.
PA (ELLE/) ELLERMAN K.
PA (STONE/) STONE D J.
PA (GROS/) GROSSE W M.
PA (ALSO/) ALSOBROOK J P.
PA (LEPL/) LEPLLEY D M.
PA (RIEG/) RIEGER D K.
PA (BURG/) BURGESS C E.
PA (EDIN/) EDINGER S R.
PA (VOSS/) VOSS E Z.
PA (MILL/) MILLER C E.

XX Mezes PD, Rastelli L, Herrmann JL, Macdougall JR, Zhong H;
PI Casman SJ, Boldog FL, Shimkets RA, Gorman L, Eisen AJ, Spaderna SK;
PI Vernet CM, Bergs C, Spytek KA, Dipippo VA, Zernusen BD, Peyman JA;
PI Ellerman K, Stone DJ, Grose WM, Alsbrook JP, Lepley DM, Rieger DK;
PI Burgess CE, Edinger SR, Voss EZ, Miller CE;
XX
XX WPI; 2004-130990/13.
DR P-PSDB; ADL56830.

XX Novel isolated SECL and/or NOXY polypeptide, useful for treating cancer,
PT cardiovascular disease, psoriasis, wound healing, and stroke.

XX Claim 24; SEQ ID NO 39; 315pp; English.

XX The invention relates to human SECC and NOXY polypeptides and the
CC polynucleotides encoding them. The invention also relates to antibodies
CC that bind immunospecifically to the polypeptides. The sequences are
CC useful for treating or preventing angiogenic-associated disorders,
CC cancer, cardiovascular disease, psoriasis, wound healing, stroke,
CC cardiomyopathy, atherosclerosis, cell signal processing disorders,
CC metabolic pathway modulation disorders, diabetes, hyperproliferative
CC diseases, cirrhosis, keloids, psoriasis, osteoarthritis, atherosclerotic
CC plaque formation, haemorrhage, ischaemic heart disease, renal disease,
CC thrombosis, hypertension, hypothyroidism, autoimmune disorders, multiple
CC sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune
CC pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis,
CC myasthenia gravis, graft-versus-host disease, asthma, peridontal
CC disease, Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC amyotrophic lateral sclerosis and shy-draeger syndrome. This sequence
CC represents a human NOXY polynucleotide of the invention.

XX Sequence 2363 BP; 456 A; 696 C; 808 G; 403 T; 0 U; 0 Other;

Query Match 100.0%; Score 2363; DB 12; Length 2363;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGAGCCCGCTTTCAGAGGACCTTACCTGAGGCGCCACAGGTGAGGCGAGCTGAGCTTAC 60

DB 2363 TCGAGCCCGCTTTCAGAGGACCTTACCTGAGGCGCCACAGGTGAGGCGAGCTGAGCTTAC 2304
QY AGGCCCCACGCGCACCGGCTCTGCTCCAGGCGCGCGCTGTGGGGGCCACCATGCTCC 120
DB AGGCCCCACGCGCACCGGCTCTGCTCCAGGCGCGCGCTGTGGGGGCCACCATGCTCC 2244
QY TGCCGAGGCGCTGAGAGCTGAGCCGACCCGAGCACTACCTGAGGCTCCGCCCCACCTGC 180
DB TGCCGAGGCGCTGAGAGCTGAGCCGACCCGAGCACTACCTGAGGCTCCGCCCCACCTGC 2184
QY TGAGCCCGAGGAGTAAAGGACAAGGGGCCAGACTCACAGTTCCAGCCCTGAGACAGGG 240
DB TGAGCCCGAGGAGTAAAGGACAAGGGGCCAGACTCACAGTTCCAGCCCTGAGACAGGG 2124
QY TTCCCTCATCCCCCACCAGCTTAATGCCCACTCTCTAATAGGGGTTCTGGGGACC 300
DB TTCCCTCATCCCCCACCAGCTTAATGCCCACTCTCTAATAGGGGTTCTGGGGACC 2064
QY TGAAGAGGGGGACATAGAGTCTCCCAAGACCTAGTGTGTCTGCTGCTCTTCTT 360
DB TGAAGAGGGGGACATAGAGTCTCCCAAGACCTAGTGTGTCTGCTGCTCTTCTT 2004
QY CAGACTAGCCGTTGACCCCAAGTCTTTCCTCCCAAGACCCAGAGTTCCAGCCTCAG 420
DB CAGACTAGCCGTTGACCCCAAGTCTTTCCTCCCAAGACCCAGAGTTCCAGCCTCAG 2004
QY GCGCCCTCTCTCTATCTAGGAGTCTTGCCCCCAAAATTCCTCTTCCCAAGACTTA 480
DB GCGCCCTCTCTCTATCTAGGAGTCTTGCCCCCAAAATTCCTCTTCCCAAGACTTA 1943
QY GCGCCCTCTCTCTATCTAGGAGTCTTGCCCCCAAAATTCCTCTTCCCAAGACTTA 1884
DB GCGCCCTCTCTCTATCTAGGAGTCTTGCCCCCAAAATTCCTCTTCCCAAGACTTA 481
QY TGATTTAGGTTCTCAGCTGTCTCTCTCTCAAAACGGAGATCTCAGTCCCTGCTCAC 540
DB TGATTTAGGTTCTCAGCTGTCTCTCTCTCAAAACGGAGATCTCAGTCCCTGCTCAC 1883
QY CAGGCTGAGGATGAGGGGTCCCATCTCTGCAAAATCCAGGCGTCCCGCTGCTGTCA 600
DB CAGGCTGAGGATGAGGGGTCCCATCTCTGCAAAATCCAGGCGTCCCGCTGCTGTCA 1823
QY GACACTGACCCCATCTCTTGAACCCAGCCCAATCTGCTCTCTATACAGGCTGCTTG 660
DB GACACTGACCCCATCTCTTGAACCCAGCCCAATCTGCTCTCTATACAGGCTGCTTG 1763
QY CCAAGGCGGATGCTCTGAGCTGCTGAGTGAACGCTGAGGAGTGGAGGCGCAGGACT 720
DB CCAAGGCGGATGCTCTGAGCTGCTGAGTGAACGCTGAGGAGTGGAGGCGCAGGACT 1703
QY GGGCTGGGCTGGGCTCCCGAGGCGCTGCTCCCGTCCATCTCTTCCACAGGTTCCACC 780
DB GGGCTGGGCTGGGCTCCCGAGGCGCTGCTCCCGTCCATCTCTTCCACAGGTTCCACC 1643
QY TGGCCGAGGAGTCTGAGGCAAGGAAATCTTAAACAAGAGGAGTGAACATGAGGAGAG 840
DB TGGCCGAGGAGTCTGAGGCAAGGAAATCTTAAACAAGAGGAGTGAACATGAGGAGAG 1583
QY GGTGCGGAGGAGTCTGAGGCAAGGAAATCTTAAACAAGAGGAGTGAACATGAGGAGAG 841
DB GGTGCGGAGGAGTCTGAGGCAAGGAAATCTTAAACAAGAGGAGTGAACATGAGGAGAG 1523
QY GTGGCGGAGGAGTCTGAGGCAAGGAAATCTTAAACAAGAGGAGTGAACATGAGGAGAG 1523
DB GTGGCGGAGGAGTCTGAGGCAAGGAAATCTTAAACAAGAGGAGTGAACATGAGGAGAG 1464
QY CTGCTACTTCTGACAGGCAATCGGGGCGGATCTCTGAGGCGCATTTGGGCTGTCTCAGG 960
DB CTGCTACTTCTGACAGGCAATCGGGGCGGATCTCTGAGGCGCATTTGGGCTGTCTCAGG 1463
QY AGTAGACAGGAGCGCTGTATCCAGTAGCAGTACAGTCTGCGAGCGTCCATATGTC 1020
DB AGTAGACAGGAGCGCTGTATCCAGTAGCAGTACAGTCTGCGAGCGTCCATATGTC 961
QY AGTAGACAGGAGCGCTGTATCCAGTAGCAGTACAGTCTGCGAGCGTCCATATGTC 1403
DB AGTAGACAGGAGCGCTGTATCCAGTAGCAGTACAGTCTGCGAGCGTCCATATGTC 1021
QY TTGACAAAGACGGAAGGAGCTGGGCGGCTGCTGCTCTGCGCTCAACGCGAGGATA 1080
DB TTGACAAAGACGGAAGGAGCTGGGCGGCTGCTGCTCTGCGCTCAACGCGAGGATA 1343
QY GCGGACTCAGCTGCGAGGAGATGGGCTTCTTCAAGGCGACTGACCTCGAGCTGAC 1140
DB GCGGACTCAGCTGCGAGGAGATGGGCTTCTTCAAGGCGACTGACCTCGAGCTGAC 1081

D 1283 GCCGAGCTCAGCTGCGAGAGATGAGGCTTCTCAGGAGCAGTACCCACTCCGAGCTGAGC 1224
Q 1141 GTGGCAAGGGGGGGGGCCCAATGGCACTGAGGCTTCTTGTGTGTGAAGAGGGAGGCTG 1200
D 1223 GTGCAAGGGGGGGGGCCCAATGGCACTGAGGCTTCTTGTGTGTGAAGAGGGAGGCTG 1164
Q 1201 CCCACACCCAGAGGCTGTGAGGCTCACTCCGTGTGTGATTTGCCCCAGAGGCGCTTTC 1260
D 1163 CCCACACCCAGAGGCTGTGAGGCTCACTCCGTGTGTGATTTGCCCCAGAGGCGCTTTC 1104
Q 1261 TTGGCCGCATCTGCGCAAGACTGTGAGCGAGAACTGCCCCGTGAGCGCATCTGTGGA 1320
D 1103 TTGGCCGCATCTGCGCAAGACTGTGAGCGAGAACTGCCCCGTGAGCGCATCTGTGGA 1044
Q 1321 GGGCCGGGACACCAAGCTTGGAGCGGTGGCCGTGGCAAGTCAAGCTTGCCTGATGATGAGA 1380
D 1043 GGGCCGGGACACCAAGCTTGGAGCGGTGGCCGTGGCAAGTCAAGCTTGCCTGATGATGAGA 984
Q 1381 CACCTCTGTGGGGGAGTCCCTGCTCCGGGGAGCTGGGTGTGAGACGGCGCCCACTGCTTC 1440
D 983 CACCTCTGTGGGGGAGTCCCTGCTCCGGGGAGCTGGGTGTGAGACGGCGCCCACTGCTTC 924
Q 1441 CCGAGCGGAAACCGGGTCTGTCCCGATGCGAGTGTGGCCGTGCGGTGCGCCAGGCGC 1500
D 923 CCGAGCGGAAACCGGGTCTGTCCCGATGCGAGTGTGGCCGTGCGGTGCGCCAGGCGC 864
Q 1501 TCTCCCAACGGTCTGCACTGAGGAGGTGAGGCTGTGTCTACACAGGAGGCTATTTCCC 1560
D 863 TCTCCCAACGGTCTGCACTGAGGAGGTGAGGCTGTGTCTACACAGGAGGCTATTTCCC 804
Q 1561 TTTCGGGAGCCCAACAGGAGGAGAGAACAGACATATTCGCTGTGCTCACTCTCCAGT 1620
D 803 TTTCGGGAGCCCAACAGGAGGAGAGAACAGACATATTCGCTGTGCTCACTCTCCAGT 744
Q 1621 CCCCTGCCCCACAGAAATACATCCAGCTGTGTGCTCCAGCTGCGGCGACAGGCGCTG 1680
D 743 CCCCTGCCCCACAGAAATACATCCAGCTGTGTGCTCCAGCTGCGGCGACAGGCGCTG 684
Q 1681 GTGATGCGCAAGATTTGATCCGTGACGAGGCTGGGGCAACAGCAGTACTATGCGCAACAG 1740
D 683 GTGATGCGCAAGATTTGATCCGTGACGAGGCTGGGGCAACAGCAGTACTATGCGCAACAG 624
Q 1741 GCCGGGGTACTCCAGGAGGCTCGAGTCCCAATACACATGATGTGCAATGAGGCGCT 1800
D 623 GCCGGGGTACTCCAGGAGGCTCGAGTCCCAATACACATGATGTGCAATGAGGCGCT 564
Q 1801 GACTTCTATGGAACCAAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGGGTGGC 1860
D 563 GACTTCTATGGAACCAAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGGGTGGC 504
Q 1861 ATTGATGCTGCCAGGAGCGAGACGCGGTGCTCCCTTGTGTGTGAGAGCAGATCTTGGG 1920
D 503 ATTGATGCTGCCAGGAGCGAGACGCGGTGCTCCCTTGTGTGTGAGAGCAGATCTTGGG 444
Q 1921 ACGCAGCTTGGCGGCTGTGTGAGATTTGTGAGTTGGGCACTGGCTGCGCTGCGCCAG 1980
D 443 ACGCAGCTTGGCGGCTGTGTGAGATTTGTGAGTTGGGCACTGGCTGCGCTGCGCCAG 384
Q 1981 AAGCAGGCGCTCTACACCAAGATCAGTACTTCGCGAGTGGATCTTCCAGGCCATAAG 2040
D 383 AAGCAGGCGCTCTACACCAAGATCAGTACTTCGCGAGTGGATCTTCCAGGCCATAAG 324
Q 2041 ACTCACTCCGAAGCCAGCGGAGTGTGACCAAGCTCTGACCGGTGAGCTTCTGCTGCGCA 2100
D 323 ACTCACTCCGAAGCCAGCGGAGTGTGACCAAGCTCTGACCGGTGAGCTTCTGCTGCGCA 264
Q 2101 GCGTCCAGGGCCGAGGATCCCGGTGTGGATCCAGCGTGGGCGCAGAGATGGAGAGT 2160
D 263 GCGTCCAGGGCCGAGGATCCCGGTGTGGATCCAGCGTGGGCGCAGAGATGGAGAGT 204
Q 2161 TTTTCTTCTGGGCGCGGTCCACAGGTCCAGAGACACCTCTCCAGAGGTCTCTCTTC 2220
D 203 TTTTCTTCTGGGCGCGGTCCACAGGTCCAGAGACACCTCTCCAGAGGTCTCTCTTC 144

Q 2221 CACAGTGGGGGGCCCACTCAGCCCGGAGACCAACCAACTCAGCTCCTGACCCCGCATGT 2280
D 143 CACAGTGGGGGGCCCACTCAGCCCGGAGACCAACCAACTCAGCTCCTGACCCCGCATGT 84
Q 2281 AAATATTTCTGCTGTGTGGAGACTCTGTCTAGAGTCCCTGATGATGAGATGCTCTT 2340
D 83 AAATATTTCTGCTGTGTGGAGACTCTGTCTAGAGTCCCTGATGATGAGATGCTCTT 24
Q 2341 AAATATTAAGATGGTTTGATT 2363
D 23 AAATATTAAGATGGTTTGATT 1
RESULT 7
AD039213/c
ID AD039213 standard; cDNA; 2363 BP.
XX
AC AD039213;
XX
DT 15-JUL-2004 (first entry)
XX
DE Human cDNA encoding novel protein NOV8.
XX
KW Human; ss; gene; NOVX; cardiomyopathy; atherosclerosis; hypertension;
KW congenital heart defect; aortic stenosis; atrial septal defect;
KW atrioventricular canal defect; ductus arteriosus; pulmonary stenosis;
KW subaortic stenosis; ventricular septal defect; valve disease;
KW tuberosus sclerosis; scleroderma; obesity; transplantation;
KW congenital adrenal hyperplasia; prostate cancer; neoplasm;
KW adenocarcinoma; lymphoma; uterus cancer; hemophilia; hypercoagulation;
KW idiopathic thrombocytopenic purpura; immunodeficiency;
KW giant versus host disease; AIDS; bronchial asthma; Crohn's disease;
KW multiple sclerosis; angiogenic-associated disorder; psoriasis;
wound healing; stroke.
XX
OS Homo sapiens.
XX
PN US2004018196-A1.
XX
PD 29-JAN-2004.
XX
PE 11-JAN-2002; 2002US-00044564.
XX
PR 11-JAN-2001; 2001US-0261013P.
XX
PR 11-JAN-2001; 2001US-0261014P.
XX
PR 11-JAN-2001; 2001US-0261018P.
XX
PR 11-JAN-2001; 2001US-0261026P.
XX
PR 11-JAN-2001; 2001US-0261029P.
XX
PR 17-AUG-2001; 2001US-0313170P.
XX
PR 10-SEP-2001; 2001US-0318410P.
XX
PA (MEZE/) MEZES P D.
PA (RAST/) RASTELLI L.
PA (HERR/) HERRMANN J L.
PA (MACD/) MACDOUGALL J R.
PA (ZHON/) ZHONG H.
PA (CASW/) CASMAN S J.
PA (BOLD/) BOLDOG F L.
PA (SHIM/) SHIMKETS R A.
PA (GORM/) GORMAN L.
PA (EISE/) EISEN A J.
PA (SPAD/) SPADERNA S K.
PA (VERN/) VERNET C A M.
PA (BERG/) BERGHS C.
PA (SPYR/) SPYTEK K A.
PA (DIPR/) DIPRO V A.
PA (ZERH/) ZERHUSEN B D.
PA (PEYM/) PEYMAN J A.
PA (ELIE/) ELIERMAN K.
PA (STON/) STONE D J.
PA (GROS/) GROSSE W M.
PA (ALSO/) ALSOBROOK J P.

PA (LEPL/) LEPLLEY D M.
 PA (BURG/) RIEGER D K.
 PA (BURG/) BURGESS C E.
 PA (EDIN/) EDINGER S R.
 PI Meares PD, Rastelli L, Herrmann JL, Macdougall JR, Zhong H,
 PI Casman SJ, Boldog FL, Shinkets RA, Gorman L, Eisen AJ, Spaderma SK,
 PI Vernet CAM, Bergs C, Spletter KA, Dipippo VA, Zernusen BD, Payman JA,
 PI Ellerman K, Stone DJ, Grose WM, Alsdorff JP, Lepley DM, Rieger DK,
 PI Burgess CE, Edinger SR,
 XX WPI: 2004-121988/12.
 DR P-PSDB; ADO39214.
 XX
 PT New isolated SECK polypeptides and polynucleotides, useful for treating
 PT or preventing, e.g. cardiomyopathy, atherosclerosis, obesity, prostate
 PT cancer, hemophilia, graft versus host disease, Crohn's disease, multiple
 PT sclerosis or psoriasis.
 PS
 PS Claim 23; SEQ ID NO 39; 306bp; English.

XX The invention relates to a method of treating or delaying the onset of an
 CC angiogenic-associated disorder comprising administering an antibody to
 CC the polypeptide of SECK. The CDNA and proteins for SECK-SECK12 are
 CC disclosed as new. Also included are an isolated NOVX (NOV1-NOV8)
 CC polypeptide (or its mature form or variant), the encoding nucleic acid
 CC for the NOVX protein described above, a vector comprising the nucleic
 CC acid molecule, a cell comprising the vector, an antibody that binds
 CC immunospecifically to NOVX, a method of treating or delaying the onset of
 CC an angiogenic-associated disorder, a method for determining the presence
 CC of or predisposition to a disease associated with altered levels of SECK1
 CC (or altered levels of the nucleic acid of SECK1 in a first mammalian
 CC subject), a method for determining the presence of or predisposition to a
 CC disease associated with altered levels of SECK1-SECK12, a method of
 CC treating a pathological state in a mammal, a method of treating or
 CC delaying the onset of a disorder, a method for determining the presence
 CC or amount of the polypeptide or nucleic acid, a method of identifying an
 CC agent that binds to a polypeptide, a method for identifying an agent that
 CC modulates the expression or activity of the polypeptide, a method for
 CC modulating the activity of the polypeptide, a method of treating or
 CC preventing a SECK-associated or NOVX disorder, a pharmaceutical
 CC composition (comprising NOVX or SECK, the nucleic acid molecule or the
 CC antibody and a pharmaceutical carrier), a kit comprising (in one or more
 CC containers) the pharmaceutical composition and a method for determining
 CC the presence of or predisposition to a disease associated with altered
 CC levels of the polypeptide or the nucleic acid in a first mammalian
 CC subject. The SECK/NOVX polypeptides and polynucleotides are useful for
 CC cardiomyopathy, atherosclerosis, hypertension, congenital heart defects,
 CC aortic stenosis, atrial septal defect, atrioventricular canal defect,
 CC ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular
 CC septal defect, valve disease, tuberosus sclerosis, scleroderma, obesity,
 CC transplantation, congenital adrenal hyperplasia, prostate cancer,
 CC neoplasm, adenocarcinoma, lymphoma, uterine cancer, haemophilia,
 CC hypercoagulability, idiopathic thrombocytopenic purpura,
 CC immunodeficiencies, graft versus host disease, AIDS, bronchial asthma,
 CC Crohn's disease or multiple sclerosis. They are also useful for treating
 CC angiogenic-associated disorders including psoriasis, wound healing or
 CC stroke. The present sequence encodes a NOVX protein.
 XX

Seq Sequence 2363 BP; 456 A; 696 C; 808 G; 403 T; 0 U; 0 Other;

Query Match 100.0%; Score 2363; DB 12; Length 2363;
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;
 Matches 2363; Conservative 0; Indels 0;

QY 1 TCGAGCCGCTTTCCAGGAGCCTTACCTGAGGAGCCACAGGCTGAGGAGCTGAGC 60
 Db 2363 TCGAGCCGCTTTCCAGGAGCCTTACCTGAGGAGCCACAGGCTGAGGAGCTGAGC 2304
 QY 61 AGGCCCCAGCCAGCCGCTTCTGCTTCCAGGCGCGCCGCTGCTGGGGGACCAATGCTCC 120
 Db 2303 AGGCCCCAGCCAGCCGCTTCTGCTTCCAGGCGCGCCGCTGCTGGGGGACCAATGCTCC 2244

QY 121 TGCCAGAGCCTTGAGAGCTGAGCCGAGCCCGGACCTACCTGAGAGCTCCGCCACCTGC 180
 Db 2243 TGCCAGAGCCTTGAGAGCTGAGCCGAGCCCGGACCTACCTGAGAGCTCCGCCACCTGC 2184
 QY 181 TGAGCCAGGAGTAAAGGAGCAAGAGGCGCCGAGACTCAGAGTTCCAGCCCTGAGAGCAGAGG 240
 Db 2183 TGAGCCAGGAGTAAAGGAGCAAGAGGCGCCGAGACTCAGAGTTCCAGCCCTGAGAGCAGAGG 2124
 QY 241 TTCCCTCATCCCCCAGCCAGCCCTTAATAGCCCACTCTTAATAGAGGGGTTCTGGGGAC 300
 Db 2123 TTCCCTCATCCCCCAGCCAGCCCTTAATAGCCCACTCTTAATAGAGGGGTTCTGGGGAC 2064
 QY 301 TGAAGAGGGGACATATGAGTCTCTCCCAAGACCTAGTGTCTCTCTCTCTCTCTCT 360
 Db 2063 TGAAGAGGGGACATATGAGTCTCTCCCAAGACCTAGTGTCTCTCTCTCTCTCTCT 2004
 QY 361 CAGACTAGCCGCTTGAAGCCAGCCAGCTCTTCTCTCCAGACCCAGAGTTCCAGCCCTCAG 420
 Db 2003 CAGACTAGCCGCTTGAAGCCAGCCAGCTCTTCTCTCCAGACCCAGAGTTCCAGCCCTCAG 1944
 QY 421 GCCCT 480
 Db 1943 GCCCT 1884
 QY 481 TGAATTGAGGTCCTCAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
 Db 1883 TGAATTGAGGTCCTCAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1824
 QY 541 CAGGCTCAGGATGAGGGGCTCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
 Db 1823 CAGGCTCAGGATGAGGGGCTCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1764
 QY 601 GACACTGACCCCATCT 660
 Db 1763 GACACTGACCCCATCT 1704
 QY 661 CCAAGGCGGAGTCCCTGACCT 720
 Db 1703 CCAAGGCGGAGTCCCTGACCT 1644
 QY 721 GGGCTGGGCTGGGCTCCCGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
 Db 1643 GGGCTGGGCTGGGCTCCCGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1584
 QY 781 TGGCCAGAGGCTCAGCCAGGAGTCAATTAAAGAGGAGCTGACATGGGCGAGAGAGAG 840
 Db 1583 TGGCCAGAGGCTCAGCCAGGAGTCAATTAAAGAGGAGCTGACATGGGCGAGAGAGAG 1524
 QY 841 GGTGGCCGAGTGTGCT 900
 Db 1523 GGTGGCCGAGTGTGCT 1464
 QY 901 CTGCTACTTGTGACAGCCATCGGAGGCGGATCTCTCTCTCTCTCTCTCTCTCTCTCT 960
 Db 1463 CTGCTACTTGTGACAGCCATCGGAGGCGGATCTCTCTCTCTCTCTCTCTCTCTCTCT 1404
 QY 961 AGTGAACAGAGAGCGCTGTATCCAGTCCAGTCAAGTCTCTCTCTCTCTCTCTCTCTCT 1020
 Db 1403 AGTGAACAGAGAGCGCTGTATCCAGTCCAGTCAAGTCTCTCTCTCTCTCTCTCTCT 1344
 QY 1021 TTGACAGAGAGAGAGAGAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
 Db 1343 TTGACAGAGAGAGAGAGAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1284
 QY 1081 GCCGAGTCTAGCTGAGAGAGATGGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
 Db 1283 GCCGAGTCTAGCTGAGAGAGATGGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1224
 QY 1141 GTGGAACCGGCGGCGGAGTGTGCACTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
 Db 1223 GTGGAACCGGCGGCGGAGTGTGCACTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1164
 QY 1201 CCCACAGCCAGAGAGCTGTGAGAGTCACTCTGCTGTGATTTGCCAGAGGCGCTTTC 1260

Db 1163 CCCACACCCAGAGGCTGTGAGGTCATCTCCGTGTGATGTGCCCCAGAGGCCGTTTC 1104
Qy 1261 TTGGCCGCGCATCTGCGCAAGACTGTGGCCGAGAAAGCTGCCGTGGAGCCGATCGTGGGA 1320
Db 1103 TTGGCCGCGCATCTGCGCAAGACTGTGGCCGAGAAAGCTGCCGTGGAGCCGATCGTGGGA 1044
Qy 1321 GGGCCGGGACACCAAGTTGGGCGGCTGGGCGGTGGCAAGTCAGCTTTCGCTATGATGAGACA 1380
Db 1043 GGGCCGGGACACCAAGTTGGGCGGCTGGGCGGTGGCAAGTCAGCTTTCGCTATGATGAGACA 984
Qy 1381 CACCTCTGTGGGGGATCCCTGCTCCGGGGACGTGGGCTGACAGCCGCCCATGCTTC 1440
Db 983 CACCTCTGTGGGGGATCCCTGCTCCGGGGACGTGGGCTGACAGCCGCCCATGCTTC 924
Qy 1441 CCGAGCGGAAACCGGGCTCTGTCCCGATGGCGAGTGTGGCCGCTGCGGCGCCAGCC 1500
Db 923 CCGAGCGGAAACCGGGCTCTGTCCCGATGGCGAGTGTGGCCGCTGCGGCGCCAGCC 864
Qy 1501 TCTCCCGACGGTCTGACACTGGGGGTGACAGCTGTGTCTACACAGGGGGCTATCTTCC 1560
Db 863 TCTCCCGACGGTCTGACACTGGGGGTGACAGCTGTGTCTACACAGGGGGCTATCTTCC 804
Qy 1561 TTTGGGAGCCCAACAGGAGGAGAAACAGACATATTGGCCCTGGTCCACTCTCCAGT 1620
Db 803 TTTGGGAGCCCAACAGGAGGAGAAACAGACATATTGGCCCTGGTCCACTCTCCAGT 744
Qy 1621 CCCCTGCCCTCAGAGAAATACATCCAGCTGTGTGCTCCAGCTGCGGCGAGCCCTG 1680
Db 743 CCCCTGCCCTCAGAGAAATACATCCAGCTGTGTGCTCCAGCTGCGGCGAGCCCTG 684
Qy 1681 GTGATGGCAAGATCTGTACCTGTACGCGGCTGGGGCAACAGCAGTACTATGGCCACAG 1740
Db 683 GTGATGGCAAGATCTGTACCTGTACGCGGCTGGGGCAACAGCAGTACTATGGCCACAG 624
Qy 1741 GCGGGGGATCTCCAGAGGCTCGAGTCCCAATACAGCANATGTGTGAATGGGCT 1800
Db 623 GCGGGGGATCTCCAGAGGCTCGAGTCCCAATACAGCANATGTGTGAATGGGCT 564
Qy 1801 GACTTCTATGAAACAGATCAAGCCCAAGATGTTCTGTGTGCTACCCGAGGGTGGC 1860
Db 563 GACTTCTATGAAACAGATCAAGCCCAAGATGTTCTGTGTGCTACCCGAGGGTGGC 504
Qy 1861 ATTGATGCTTCCAGGCGGACAGCGGTGTCTCTTGTGTGTAGAGACAGATCTTCGG 1920
Db 503 ATTGATGCTTCCAGGCGGACAGCGGTGTCTCTTGTGTGTAGAGACAGATCTTCGG 444
Qy 1921 ACGCAGGTTGGCGGCTGTGTGATTTGATTTGGGCACTGGCTGTGCCCTGGCCAG 1980
Db 443 ACGCAGGTTGGCGGCTGTGTGATTTGATTTGGGCACTGGCTGTGCCCTGGCCAG 384
Qy 1981 AAGCAGGCGCTCTACACCAAAAGTCACTTCCGGAGTGGATCTTCCAGGCCATAAG 2040
Db 383 AAGCAGGCGCTCTACACCAAAAGTCACTTCCGGAGTGGATCTTCCAGGCCATAAG 324
Qy 2041 ACTCACTCCGAAGCCAGCGGATGTGAACCAAGCTCTGACCGGTGTCTTCGCTGGCA 2100
Db 323 ACTCACTCCGAAGCCAGCGGATGTGAACCAAGCTCTGACCGGTGTCTTCGCTGGCA 264
Qy 2101 GCGTCCAGGGGCGAGGATGATCCGGGTGTGGATCCAGGTCGGGCGAGGATGGGAGCT 2160
Db 263 GCGTCCAGGGGCGAGGATGATCCGGGTGTGGATCCAGGTCGGGCGAGGATGGGAGCT 204
Qy 2161 TTTTCTTCTTGGGCGGCTCCAGAGTCCAGAGACACCTCTCCAGAGGCTCTCTTC 2220
Db 203 TTTTCTTCTTGGGCGGCTCCAGAGTCCAGAGACACCTCTCCAGAGGCTCTCTTC 144
Qy 2221 CACAGTGGCGGCGCACTCAGCCCGAGAGACCAACCACTCAACCTCTCTGACCCCATGT 2280
Db 143 CACAGTGGCGGCGCACTCAGCCCGAGAGACCAACCACTCAACCTCTCTGACCCCATGT 84
Qy 2281 AAATATTGTTGCTGTGGAGTCCCTGTAGAGTGCCTGTAGATGGATGCTCTT 2340

Db 83 AAATATTGTTGCTGTGGAGTCCCTGTAGAGTGCCTGTAGATGGAGTGTCTT 24
Qy 2341 AAATATTAAAGATGTTTGAAT 2363
Db 23 AAATATTAAAGATGTTTGAAT 1
RESULT 8
ACC46451
ID ACC46451 standard; cDNA; 2311 BP.
XX
XX ACC46451;
AC 02-JUN-2003 (first entry)
XX
XX Human dlthp protein modification/maintenance protein-encoding cDNA.
DE Human; dlthp; diagnostic and therapeutic polynucleotide; diagnosis;
XX cancer; cell proliferative disorder; autoimmune disorder;
KW cancer; cell proliferative disorder; autoimmune disorder; diagnosis;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW protein modification; protein maintenance; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200297031-A2.
PN
XX
XX 05-DEC-2002.
PD
XX
XX 27-MAR-2002; 2002MO-US010056.
PE
XX
XX 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291282P.
PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuson O, Yap PE, Ameshey SR;
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleeheid Y, Gierlein EH;
PI Peraltia CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B,
PI Flores V, Marwaha R, Lo A, Lan RV, Urashka ME;
XX
XX WPI: 2003-129518/12.
DR P-PSDB: ABR41513.
XX
XX
XX Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
XX Claim 2; SEQ ID NO 372; 591bp; English.
PS
XX The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dlthp (ACC46080-ACC46749) and to their encoded
CC proteins (dlthp; ABR41136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dlthp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dlthp nucleic acid sequence; the
CC recombinant production of dlthp proteins; antibodies specific for dlthp
CC proteins; microarrays comprising dlthp nucleic acid sequences; methods of
CC detecting dlthp polynucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a dlthp protein; and methods of
CC assessing the toxicity of test compounds using a dlthp hybridisation
CC probe. Dithp nucleic acid sequences and dlthp proteins may be used in the

CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; hormonal disorders; metabolic,
CC disorders; neurological disorders; gastrointestinal disorders; transport
CC disorders; and connective tissue disorders. They may also be used to
CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock
CC in humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a dithp cDNA encoding a DITHP protein
CC which is involved in protein modification and/or maintenance. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX

Sequence 2311 BP; 399 A; 786 C; 678 G; 448 T; 0 U; 0 Other;

Query Match 97.5%; Score 2303.4; DB 8; Length 2311;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 59 GCAGGCCCCACGCGCCTCTGCTCCAGCGCGCGCTGTCGGGGGCACCATGCT 118
Db 1 GCAGGCCCCACGCGCCTCTGCTCCAGCGCGCGCTGTCGGGGGCACCATGCT 60
Qy 119 CTTGCCCAAGGCTTGAAGTACGACCGGACCTAATCTGAGGCTCGGCCCACT 178
Db 61 CCTGCCCAAGGCTTGAAGTACGACCGGACCTAATCTGAGGCTCGGCCCACT 120
Qy 179 GCTGGACCCCAAGGCTTGAAGTACGACCGGACCTAATCTGAGGCTCGGCCCACT 238
Db 121 GCTGGACCCCAAGGCTTGAAGTACGACCGGACCTAATCTGAGGCTCGGCCCACT 180
Qy 239 GGTTCCTCCATCCCGCCACCGAGCTAATGCGCACTCTAATAGAGGGTTCCTGGGGA 298
Db 181 GGTTCCTCCATCCCGCCACCGAGCTAATGCGCACTCTAATAGAGGGTTCCTGGGGA 240
Qy 299 CTTGAAGAGGGGGGACTATGAGCTCTCCCAAGCACTAAGTGTCTGCTCTCTCC 358
Db 241 CCTGAAGAGGGGGGACTATGAGCTCTCCCAAGCACTAAGTGTCTGCTCTCTCC 300
Qy 359 TTTGAGACTCAAGCGGTGGAACCCCAAGTCTTTCTCCCAAGCACTAAGTGTCTGCTCTC 418
Db 301 TTTGAGACTCAAGCGGTGGAACCCCAAGTCTTTCTCCCAAGCACTAAGTGTCTGCTCTC 360
Qy 419 AGGCCCTCTCTCCCTCATAGGAGTCTGCGCCCAATTTCTCTTCCCAAGACT 478
Db 361 AGGCCCTCTCTCCCTCATAGGAGTCTGCGCCCAATTTCTCTTCCCAAGACT 420
Qy 479 TATGATTTCAAGGTCTCTGAGTGTCTCTCTCTCAACCGGGATCTCTAGTCCCTGCTCC 538
Db 421 TATGATTTCAAGGTCTCTGAGTGTCTCTCTCTCAACCGGGATCTCTAGTCCCTGCTCC 480
Qy 539 ACCAGGCTCAAGGCAATGGGGGTCTCCCATCTCTGCAATCAAGGCTCTCCCGCTGCTGCT 598
Db 481 ACCAGGCTCAAGGCAATGGGGGTCTCCCATCTCTGCAATCAAGGCTCTCCCGCTGCTGCT 540
Qy 599 CAGACACTGACCCCATCTTGAACCGAGCCCAATCTGCGCTCGTATCAAGGCGTGCCT 658
Db 541 CAGACACTGACCCCATCTTGAACCGAGCCCAATCTGCGCTCGTATCAAGGCGTGCCT 600
Qy 659 GGCCCAAGGCCAGTCTCTTACAGCTGCTGATGAGACGCTGGAGCTGGGGGCGCCAGAGA 718
Db 601 GGCCCAAGGCCAGTCTCTTACAGCTGCTGATGAGACGCTGGAGCTGGGGGCGCCAGAGA 660
Qy 719 CTGGGCTGGGCTGGGCTCTCCCAAGGCTGCTCTCCATCTCTCAAGGTTCCAC 778
Db 661 CTGGGCTGGGCTGGGCTCTCCCAAGGCTGCTCTCCATCTCTCAAGGTTCCAC 720

Qy 779 CTTGGCCCAAGAGTCAAGCCAGGGAATCATTTAAACAAGGCAAGTACATGCGCAGAGG 838
Db 721 CTTGGCCCAAGAGTCAAGCCAGGGAATCATTTAAACAAGGCAAGTACATGCGCAGAGG 780
Qy 839 AGGGTGGCCGAGTCTGTCATGCTGCTCCAGACCCAAAGTGGCAGCTCTCACTGCGGGA 898
Db 781 AGGGTGGCCGAGTCTGTCATGCTGCTCCAGACCCAAAGTGGCAGCTCTCACTGCGGGA 840
Qy 899 CCCCTGACTCTTGAACGCAATGGGGGGGCAATCTGGGGGCAATGGGCTGTTCTCTCA 958
Db 841 CCCCTGACTCTTGAACGCAATGGGGGGGCAATCTGGGGGCAATGGGCTGTTCTCTCA 900
Qy 959 GGAATGACCAAGAGCCGCTTACCAAGTGAAGTCACTCTGCGGAGCGCTCGCTCAATGG 1018
Db 901 GGAATGACCAAGAGCCGCTTACCAAGTGAAGTCACTCTGCGGAGCGCTCGCTCAATGG 960
Qy 1019 TCTTTGACAAAGCGAAAGGAGCGTGGCGCTGCTGCTCTCTGCGCTCAAGCCAGGG 1078
Db 961 TCTTTGACAAAGCGAAAGGAGCGTGGCGCTGCTGCTCTCTGCGCTCAAGCCAGGG 1020
Qy 1079 TAGCCGGAATCAAGTGGAGAGATGGGGCTTCTCAAGGGGACATGACCTCGGAGCTGG 1138
Db 1021 TAGCCGGAATCAAGTGGAGAGATGGGGCTTCTCAAGGGGACATGACCTCGGAGCTGG 1080
Qy 1139 ACGTGCGAAGCGCGCGCCCAATGGCACTGCGGCTTCTTCTGTGGAAGAGGGGAGGC 1198
Db 1081 ACGTGCGAAGCGCGCGCCCAATGGCACTGCGGCTTCTTCTGTGGAAGAGGGGAGGC 1140
Qy 1199 TGCCCAACCCAGAGGCTGCTGAGTCAATCTCGTGTGATTTGCCCAAGGCGGTT 1258
Db 1141 TGCCCAACCCAGAGGCTGCTGAGTCAATCTCGTGTGATTTGCCCAAGGCGGTT 1200
Qy 1259 TCTTTGCGCGCATCTGCAAGACTGTGGCGCGCAAGAGCTGCCGCTGGAACGCACTGTTG 1318
Db 1201 TCTTTGCGCGCATCTGCAAGACTGTGGCGCGCAAGAGCTGCCGCTGGAACGCACTGTTG 1260
Qy 1319 GAGGCGGAGACACGAGCTTGGCGGCTGAGCGCAAGTCAAGCTTCTGATGATGAG 1378
Db 1261 GAGGCGGAGACACGAGCTTGGCGGCTGAGCGCAAGTCAAGCTTCTGATGATGAG 1320
Qy 1379 CACACTCTGTGGGGATCTCTGCTCTCGGGGACTGGGTCTGACAGCGGCCACTGCT 1438
Db 1321 CACACTCTGTGGGGATCTCTGCTCTCGGGGACTGGGTCTGACAGCGGCCACTGCT 1380
Qy 1439 TCCCGAGCGGAAACCGGGTCTGTGCCGATGGGAGTGTGGCGGGCGCGGAGG 1498
Db 1381 TCCCGAGCGGAAACCGGGTCTGTGCCGATGGGAGTGTGGCGGGCGCGGAGG 1440
Qy 1499 CTTCTCCCAAGGCTTGAAGTGGAGTGGAGCTGTGATCAACAGGGGACTATCTTC 1558
Db 1441 CTTCTCCCAAGGCTTGAAGTGGAGTGGAGCTGTGATCAACAGGGGACTATCTTC 1500
Qy 1559 CTTTGGGGAACCCCAAGCGGAGAGAAACGAAACGATATTGCGCTGTGCTCACTCTCCA 1618
Db 1501 CTTTGGGGAACCCCAAGCGGAGAGAAACGAAACGATATTGCGCTGTGCTCACTCTCCA 1560
Qy 1619 GTCCCGTCCCTCAAGAAATACATCCAGCTGTGTGCTCCCAAGTGGCGGCGGAGGCC 1678
Db 1561 GTCCCGTCCCTCAAGAAATACATCCAGCTGTGTGCTCCCAAGTGGCGGCGGAGGCC 1620
Qy 1679 TGTGTGATGGCAAGATCTGATCGTGAAGGCTGGGCAACAGCACTATGATGGCAAC 1738
Db 1621 TGTGTGATGGCAAGATCTGATCGTGAAGGCTGGGCAACAGCACTATGATGGCAAC 1680
Qy 1739 AGCGGGGGATCTTCAAGAGGCTTGAAGTCCCATTAATCAAGATGATGTCTGCAATGGCG 1798
Db 1681 AGCGGGGGATCTTCAAGAGGCTTGAAGTCCCATTAATCAAGATGATGTCTGCAATGGCG 1740
Qy 1799 CTGACTTTATGGAAGCAAGATCAAGCCCAAGATGTTCTGTGCTGCTTACCCGAGGGGTG 1858
Db 1741 CTGACTTTATGGAAGCAAGATCAAGCCCAAGATGTTCTGTGCTGCTTACCCGAGGGGTG 1800
Qy 1859 GCAATGATGCTGCGCAAGGCGAAGCGGTGCTCTTTGTGTGAGAGACGACATCTCTC 1918

Db 1801 GCATTGATGCTGCGCAGGCGCAGCGGTGCTCCCTTTGTGTGAGGACAGCATCTCTC 1860
Qy 1919 GAGCGCCACGTTGGCGGCTGTGTGGCATTTGTAGTGGGAGCACTGGTCTGCGCC 1978
Db 1861 GGACGCCACGTTGGCGGCTGTGTGGCATTTGTAGTGGGAGCACTGGTCTGCGCC 1920
Qy 1979 AGAAGCCAGGCGGTCTACACCAAGTCATGACTTCCGGAGTGGATCTTCCAGGCAATAA 2038
Db 1921 AGAAGCCAGGCGGTCTACACCAAGTCATGACTTCCGGAGTGGATCTTCCAGGCAATAA 1960
Qy 2039 AGACTCACTCCGAAGCCAGCGGAGCATGTGAGTCCAGCTGACCGGTGCTTCTGCTGCG 2098
Db 1981 AGACTCACTCCGAAGCCAGCGGAGCATGTGAGTCCAGCTGACCGGTGCTTCTGCTGCG 2040
Qy 2099 CAGCCTCCAGGCGCCGAGGTGATCCCGGTGTGGATCCAGCTGAGGCGGAGATGGAGAC 2158
Db 2041 CAGCCTCCAGGCGCCGAGGTGATCCCGGTGTGGATCCAGCTGAGGCGGAGATGGAGAC 2100
Qy 2159 GTTTTCTTCTTGGGCGCGGTCCACAGTCCAGGACACCTCTCTCCAGGCTCTCTCT 2218
Db 2101 GTTTTCTTCTTGGGCGCGGTCCACAGTCCAGGACACCTCTCTCCAGGCTCTCTCT 2160
Qy 2219 TCCACAGTGGGCGGCGCCACTCAGCGCCGAGACACCCCACTCAACCTCTGACCCCAT 2278
Db 2161 TCCACAGTGGGCGGCGCCACTCAGCGCCGAGACACCCCACTCAACCTCTGACCCCAT 2220
Qy 2279 GTAATATATTTGTGTGTCTGTGGAGCTCTGTAGTGGTCCCTGATGATGGAGTGTCT 2338
Db 2221 GTAATATATTTGTGTGTCTGTGGAGCTCTGTAGTGGTCCCTGATGATGGAGTGTCT 2280
Qy 2339 TTAATATATTAAGATGTTTGATT 2363
Db 2281 TTAATATATTAAGATGTTTGATT 2305

RESULT 9
ACN41149
ID ACN41149 standard; cDNA; 1903 BP.
XX ACN41149;
AC ACN41149;
XX
DT 18-NOV-2004 (first entry)
DE Human diagnostic and therapeutic polynucleotide SEQ ID NO:24.
XX
XX 66; gene; gene therapy; human diagnostic and therapeutic polynucleotide;
KM dltph.
XX
OS Homo sapiens.
XX
XX WO2004023973-A2.
XX
XX 25-MAR-2004.
XX
XX 12-SEP-2003; 2003WO-US028227.
XX
XX 12-SEP-2002; 2002US-0410259P.
PR 12-SEP-2002; 2002US-0410260P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F,
PI Huthorne TA, Suchorolski MT, Altus CM, Pites SD, Elder LV,
PI Mooney EM, Deleage AM, Panesar IS, Banville SC, Reddy TP,
PI Stevens KM, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH,
PI Petralia CH, Anderson SB, Rieux P, Shen EJ, Wu MC, Stuve IL,
PI Lagace RE, Spito PA, Stewart EA, Wingrove J, Vitec UA, Kitron ES,
PI Xu Y, Kwong M, Policky JL, Hurwitz BU, Ma Y, Jackson JL, Gietzen D,
PI Patry S, Shi X, Suarez CJ,
XX
XX WPI; 2004-329368/30.
DR P-PSDB; ABM82497.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 1, Page; 190pp; English.
PS
CC The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dltph) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dltph
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dltph polynucleotide of
CC the invention. Note: The sequence data for this patent is not represented
CC in the printed specification, but was obtained in electronic format
CC directly from WPI at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ Sequence 1903 BP; 316 A; 628 C; 593 G; 366 T; 0 U; 0 Other;
Query Match 69.3%; Score 1638.6; DB 13; Length 1903;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1715; Conservative 0; Mismatches 4; Indels 54; Gaps 2;
Qy 622 CCCAGCCCAATGCTGCTGCTGATCAGGCGGTGTCTGCGCAAGGCCACTCCCTAAGC 681
Db 147 CCAAGGCCCAATGCTGCTGCTGATCAGGCGGTGTCTGCGCAAGGCCACTCCCTAAGC 206
Qy 682 CTGCGTGGATGAGAGCGCTGGAGACTGGGCGGCGCAGACTGGGCTGGGCTCCCA 741
Db 207 CTGCGTGGATGAGAGCGCTGGAGACTGGGCGGCGCAGACTGGGCTGGGCTCCCA 266
Qy 742 GGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 801
Db 267 GGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 326
Qy 802 GAATCATTAACAAGAGGAGTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 861
Db 327 GAATCATTAACAAGAGGAGTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 386
Qy 862 TGCTCCAGACCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 921
Db 387 TGCTCCAGACCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 446
Qy 922 GGGGCGGATCTGCGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 981
Db 447 GGGGCGGATCTGCGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 506
Qy 982 CCAATGAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1041
Db 507 CCAATGAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 566
Qy 1042 TGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1101
Db 567 TGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 626
Qy 1102 ATGAGGCTTCTCAAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1161
Db 627 ATGAGGCTTCTCAAGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 686
Qy 1162 GGCAGCTGCGGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1221
Db 687 GGCAGCTGCGGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 746
Qy 1222 GAGTCACTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1281

```

Db 747 GAGGTCATCTCCGTGTGTGATTTGCCCCAGAGCCGTTCTTGGCCGACATCTGCCAAGAC 806
Qy 1282 TGTGGCCGAGAAAGCTGCCCCGTGAGACCGCATCGTGGAGGCGGGAGAACCAAGCTTGGGC 1341
Db 807 TGTGGCCGAGAAAGCTGCCCCGTGAGACCGCATCGTGGAGGCGGGAGAACCAAGCTTGGGC 866
Qy 1342 CGGTGGCCGAGAAAGCTGCCCCGTGAGACCGCATCGTGGAGGCGGGAGAACCAAGCTTGGGC 1401
Db 867 CGGTGGCCGAGAAAGCTGCCCCGTGAGACCGCATCGTGGAGGCGGGAGAACCAAGCTTGGGC 926
Qy 1402 CTCTCCGGGAGACTGGGTGCTGACAGCCGACCTGCTTCCGGAGCCGAGAACCGGGTCTGTG 1461
Db 927 CTCTCCGGGAGACTGGGTGCTGACAGCCGACCTGCTTCCGGAGCCGAGAACCGGGTCTGTG 986
Qy 1462 TCCCGATGGCGAGTGTTCGCGGTGCGCGTGGCCGAGGCGCTTCCCGAGCGGTCTGACAGTG 1521
Db 987 TCCCGATGGCGAGTGTTCGCGGTGCGCGTGGCCGAGGCGCTTCCCGAGCGGTCTGACAGTG 1046
Qy 1522 GGGGTGCGAGGCTGTGTGTACACAGGGGGCTATCTTCCCTTCCGGAGCCGAGAACCGGGTCTGTG 1581
Db 1047 GGGGTGCGAGGCTGTGTGTACACAGGGGGCTATCTTCCCTTCCGGAGCCGAGAACCGGGTCTGTG 1106
Qy 1582 GAGAACAGCAACGATATTGCTGCTGCTGACCTTCCAGTCTCCCTGCTTCCAGCAAGATAC 1641
Db 1107 GAGAACAGCAACGATATTGCTGCTGCTGACCTTCCAGTCTCCCTGCTTCCAGCAAGATAC 1166
Qy 1642 ATCCAGCCTGTGTGCTTCCCGAGCTGCGCGGCGAGGCGCTGTGTGAGTGGCAAGATCTGTACC 1701
Db 1167 ATCCAGCCTGTGTGCTTCCCGAGCTGCGCGGCGAGGCGCTGTGTGAGTGGCAAGATCTGTACC 1226
Qy 1702 GTGACGGGCTGGGGCAACAGCAGTACTACTAGGCCAAGCGCGGGGTACTCAGAGAGGCT 1761
Db 1227 GTGACGGGCTGGGGCAACAGCAGTACTACTAGGCCAAGCGCGGGGTACTCAGAGAGGCT 1283
Qy 1762 CGAGTCCCCATTAATGACGATGATGTCTGCAATGAGCGCTGACTTCTATGAGAACCAAGATC 1821
Db 1264 CGAGTCCCCATTAATGACGATGATGTCTGCAATGAGCGCTGACTTCTATGAGAACCAAGATC 1323
Qy 1822 AAGGCCAAGATGTCTGTGCTGCTGAGTCCCGAGGGGTGCAATGAGTGGCTGCAAGGGGAC 1881
Db 1324 AAGGCCAAGATGTCTGTGCTGCTGAGTCCCGAGGGGTGCAATGAGTGGCTGCAAGGGGAC 1383
Qy 1882 AGCGGTGCTCCCTTGTGTGTGAGAGACAGCATCTCTCGAAGCGCACGTTGGCGCTGTGT 1941
Db 1384 AGCGGTGCTCCCTTGTGTGTGAGAGACAGCATCTCTCGAAGCGCACGTTGGCGCTGTGT 1443
Qy 1942 GGCATTTGAGTGGGGGACAGTGGCTGTGCTGCGCCGAGAACCGAGCGGTGTACACCAAA 2001
Db 1444 GGCATTTGAGTGGGGGACAGTGGCTGTGCTGCGCCGAGAACCGAGCGGTGTACACCAAA 1503
Qy 2002 GTCAGTACTTCCCGGAGTGTGATCTTCCAGGCCATTA----- 2038
Db 1504 GTCAGTACTTCCCGGAGTGTGATCTTCCAGGCCATTAAGACCTCGGGAGCCCGACGCTGT 1563
Qy 2039 -----AGACTACTTCCGAGCCAGCGGATGGTGAACCCAGCTTGAACCGGTGCTTC 2090
Db 1564 CTTTCCCGAGACTACTTCCGAGCCAGCGGATGGTGAACCCAGCTTGAACCGGTGCTTC 1623
Qy 2091 TCGCTGGGAGCTTCCAGGGCCCGAGGTGATCCCGGTGTGGGATCCACGCTGGGCGGAG 2150
Db 1624 TCGCTGGGAGCTTCCAGGGCCCGAGGTGATCCCGGTGTGGGATCCACGCTGGGCGGAG 1683
Qy 2151 GATGGGAGCTTTCTTCTTGGGCGCGGTCCAGAGTTCAGAGACCCCTCCGACGGG 2210
Db 1684 GATGGGAGCTTTCTTCTTGGGCGCGGTCCAGAGTTCAGAGACCCCTCCGACGGG 1743
Qy 2211 TCTCTCTTCCAGAGTGGCGGCGCCACTCAGCCCGAGAGACCAACCTCAGCCCTCTG 2270
Db 1744 TCTCTCTTCCAGAGTGGCGGCGCCACTCAGCCCGAGAGACCAACCTCAGCCCTCTG 1803
Qy 2271 ACCCCCATGTAATATTGTCTGTCTGTGTGGGACTCTGTGTAGGTGCGCTGTATGTG 2330
Db 1804 ACCCCCATGTAATATTGTCTGTCTGTGTGGGACTCTGTGTAGGTGCGCTGTATGTG 1863

```

```

Qy 2331 GATGCTCTTAAATTAATTAAGATGTTTGATT 2363
Db 1864 GATGCTCTTAAATTAATTAAGATGTTTGATT 1896

RESULT 10
AADI3167
ID AADI3167 standard; DNA, 1783 BP.
XX
XX AADI3167;
XX
XX 16-OCT-2001 (first entry)
XX
XX Human hepsin DNA.
XX
XX Human, transmembrane serine protease; membrane-type serine protease;
XX MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
XX lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
XX hepsin; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 246..1499
XX FT /*tag= a
XX FT /product= "Human hepsin protein"
XX
XX MO200157194-A2.
XX
XX 09-AUG-2001.
XX
XX 02-FEB-2001; 2001MO-US003471.
XX
XX 03-FEB-2000; 2000US-0179982P.
XX 18-FEB-2000; 2000US-0183542P.
XX 22-JUN-2000; 2000US-0213124P.
XX 26-JUL-2000; 2000US-0220970P.
XX 08-SEP-2000; 2000US-00657986.
XX 22-SEP-2000; 2000US-0234840P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Madison EL, Ong EO, Yeh J;
XX
XX WPI; 2001-488877/53.
XX
XX P-PSDB; AA06942.
XX
XX Novel single chain polypeptide comprising protease domain of type-II
XX membrane-type serine protease or its catalytically active portion useful
XX for treating and preventing cancer and tumor.
XX
XX
XX Disclosure; Page 246-248; 256pp; English.
XX
XX The invention relates to transmembrane serine proteases and their
XX corresponding nucleotides and the protease domain of a type-II membrane-
XX type serine protease (MTSP). MTSP is useful for identifying compounds
XX that modulate or inhibits its proteolytic activity and for formulating a
XX medicament for treating neoplastic disease. MTSP and its corresponding
XX nucleotides are useful in preventing or treating tumours or cancers such
XX as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in
XX diagnostics and in hybridisation assays. MTSP is useful as a diagnostic
XX marker for tumour development, growth and/or progression and as
XX immunogens to generate antibodies that specifically bind to it. MTSP DNA
XX is useful in a yeast two-hybrid system and in gene therapy. The present
XX sequence is a DNA encoding human hepsin protein
XX
XX Sequence 1783 BP; 304 A; 580 C; 558 G; 341 T; 0 U; 0 Other;

```

```

Query Match 67.4%; Score 1593.6; DB 4; Length 1783;
Best Local Similarity 99.1%; Pred. No. 7.2e-313;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

```

OY	736	CCCCCAGGCGCTCGCTCCCGGCTCATTCTCTTCAAGAGTCCACCTCGGCCAGAGGTCA	795
Db	157	CTTCAGAGGTCTCGCCCCCACTGCTGAGACCCCA--GGATCCCACTCTGCGCCAGAGGTCA	215
OY	796	GCCAGAGGAACTATTAAACAAGAGGAGTGAATAGGGGCAAGAGGAGGTGGCCGGACTGTG	855
Db	216	GCCAGGGAATCTTAAACAAGAGGAGTGAATAGGGGCAAGAGGAGGTGGCCGGACTGTG	275
OY	856	CCATGCTGCTCCAGACCCAGAGGTGAGCTCTCACTGCGGGGACCCTGTACTTTTGACA	915
Db	276	CCATGCTGCTCCAGACCCAGAGGTGAGCTCTCACTGCGGGGACCCTGTACTTTTGACA	335
OY	916	GCCATCGGGGCGGCAATCTTGSGCAATTGTGCTGTTCTTCTCAGAGTGAACGAGCGCG	975
Db	336	GCCATCGGGGCGGCAATCTTGSGCAATTGTGCTGTTCTTCTCAGAGTGAACGAGCGCG	395
OY	976	CTGTATCCAGTGCAGGTGAGCTGTGCGGACGCTGGGCTCATGAGTCTTTTGAACAAGCGGAA	1035
Db	396	CTGTATCCAGTGCAGGTGAGCTGTGCGGACGCTGGGCTCATGAGTCTTTTGAACAAGCGGAA	455
OY	1036	GGGAGTGGGCGGCTCTGTGCTCTCTCGGCTCAACGCGCAGGGTATGCGGACTCACTGTC	1095
Db	456	GGGAGTGGGCGGCTCTGTGCTCTCTCGGCTCAACGCGCAGGGTATGCGGACTCACTGTC	515
OY	1096	GAGAGATGGGCTTCTTCAAGGCACTGACCCACTCGAGCTTGAAGTGCAGACGGCGGGC	1155
Db	516	GAGAGATGGGCTTCTTCAAGGCACTGACCCACTCGAGCTTGAAGTGCAGACGGCGGGC	575
OY	1156	GCCATTTGGCAGTCCGGGCTTCTTCTGTGTGAACGAGGGGAGGCTGCCCAACCCAGAG	1215
Db	576	GCCATTTGGCAGTCCGGGCTTCTTCTGTGTGAACGAGGGGAGGCTGCCCAACCCAGAG	635
OY	1216	CTGCTGAGAGGTATCTCCGCTGTGATTTGGCCCAAGGCGGTTTCTTGAGCGCACTCAG	1275
Db	636	CTGCTGAGAGGTATCTCCGCTGTGATTTGGCCCAAGGCGGTTTCTTGAGCGCACTCAG	695
OY	1276	CAGAAGTGTGGCCGCGAGGAAGCTGCGCGTGAACCGCATGTGTGAAGCCGGGACACAGC	1335
Db	696	CAGAAGTGTGGCCGCGAGGAAGCTGCGCGTGAACCGCATGTGTGAAGCCGGGACACAGC	755
OY	1336	TTGGGCGCGTGGCCGTGGCAAGTCAAGCTTGGCTATGTATGAGACACACTCTGTGGGGGA	1395
Db	756	TTGGGCGCGTGGCCGTGGCAAGTCAAGCTTGGCTATGTATGAGACACACTCTGTGGGGGA	815
OY	1396	TCCCTGCTCTCCGGGAGCTGGGTGTGAACGCGCGCCACTGCTTCCGGGACGGAAACCG	1455
Db	816	TCCCTGCTCTCCGGGAGCTGGGTGTGAACGCGCGCCACTGCTTCCGGGACGGAAACCG	875
OY	1456	GTCCTGTTCGATGTGCGAGTGTTCGCCGAGTCCGATGCGCCAGAGCTCTCCCCAGGTCG	1515
Db	876	GTCCTGTTCGATGTGCGAGTGTTCGCCGAGTCCGATGCGCCAGAGCTCTCCCCAGGTCG	935
OY	1516	CAGCTGGGGGTGCAAGCTGTGTGTTCAACAAGGGGCTATCTTCCCTTTCGGAACCCCAAC	1575
Db	936	CAGCTGGGGGTGCAAGCTGTGTGTTCAACAAGGGGCTATCTTCCCTTTCGGAACCCCAAC	995
OY	1576	AGCGAGGAGAACAGCAACGATATTGCCCTGTGTCACCTTCAGTCCCTGCGCCCTTCA	1635
Db	996	AGCGAGGAGAACAGCAACGATATTGCCCTGTGTCACCTTCAGTCCCTGCGCCCTTCA	1055
OY	1636	GAATATCATCCAGCTGTGTGCTCCCAAGCTGCGGCGCAGGCGCTGTGTGATGAGCAAGTC	1695
Db	1056	GAATATCATCCAGCTGTGTGCTCCCAAGCTGCGGCGCAGGCGCTGTGTGATGAGCAAGTC	1115
OY	1696	TGTATCCGTGACGGGCTGGGGGCAACGCGAGTACTATGTGCAACAGGCGGGGTACTCAG	1755
Db	1116	TGTATCCGTGACGGGCTGGGGGCAACGCGAGTACTATGTGCAACAGGCGGGGTACTCAG	1175
OY	1756	GAGGCTGAGTCCCAATAATACGAATGATGTCTGCAATGGCGGTGATTTCTATAGGAAC	1815
Db	1176	GAGGCTGAGTCCCAATAATACGAATGATGTCTGCAATGGCGGTGATTTCTATAGGAAC	1235
OY	1816	CAGATCAAGCCCAAGATGTTCTGTGCTGGCTACCCCGAGGGTGGCAATGATGCTGCGAG	1875

Db	1236	CAGATCAAGCCCAAGATGTTCTGTGCTGAGCTAACCCGAGGGTGGCATTTGATGCTTCGACG	1295
Qy	1876	GAGCGACAGCGGTGGTCCCTTTGTGTGTGAGGACAGCATCTCTGGACGCCACGTTGGCCGG	1935
Db	1236	GAGCAGACGGGTGGTCCCTTTGTGTGTGAGGACAGCATCTCTGGACGCCACGTTGGCCGG	1355
Qy	1936	CTGTGTGCAATTGTGAGTTGGGGGCACTGGCTGTGCCCTGGGCCAGAAAGCCAGGCGTCTAC	1995
Db	1356	CTGTGTGCAATTGTGAGTTGGGGGCACTGGCTGTGCCCTGGGCCAGAAAGCCAGGCGTCTAC	1415
Qy	1996	ACCAAACTCACTGATCTTCCGGGAAGTGATCTTCCAGGCATTAAGACTCACTCCGAAGCC	2055
Db	1416	ACCAAACTCACTGATCTTCCGGGAAGTGATCTTCCAGGCATTAAGACTCACTCCGAAGCC	1475
Qy	2056	AGCGGCATGTGTGACCCAGCTGTGACCGGTGGACTTCTGGCTGGCCGACGCTCCAGGGCCGGA	2115
Db	1476	AGCGGCATGTGTGACCCAGCTGTGACCGGTGGACTTCTGGCTGGCCGACGCTCCAGGGCCGGA	1535
Qy	2116	GGTGTATCCCGGTGGTGGGATCCACGCTGGGCCGAGATGGGACGTTTTTCTTTTGAGGCC	2175
Db	1536	GGTGTATCCCGGTGGTGGGATCCACGCTGGGCCGAGATGGGACGTTTTTCTTTTGAGGCC	1595
Qy	2176	CGGTTCACAGGTCCCAAGACACCTCTCCCTCCAGGGTCTCTTCCACAGTGGCGGGCCC	2235
Db	1596	CGGTTCACAGGTCCCAAGACACCTCTCCCTCCAGGGTCTCTTCCACAGTGGCGGGCCC	1655
Qy	2236	ACTCAGCCCCGAGACACACCCAACTCACCCTCCGACGCCCCAGATGTAATTAATTTCTGCT	2295
Db	1656	ACTCAGCCCCGAGACACACCCAACTCACCCTCCGACGCCCCAGATGTAATTAATTTCTGCT	1715
Qy	2296	GTCGTGGAGCTCTCTCTAGAGTGGCCCTGATGATGGATGCTCTTAAATAATAAGATGG	2355
Db	1716	GTCGTGGAGCTCTCTCTAGAGTGGCCCTGATGATGGATGCTCTTAAATAATAAGATGG	1775
Qy	2356	TTTTTGATT 2363	
Db	1776	TTTTGATT 1783	
RESULT 11			
AAS43103			
ID	AAS43103	standard; cDNA; 1783 BP.	
AC	AAS43103;		
XX			
DT	19-NOV-2002	(first entry)	
XX			
DE	cDNA encoding Hepsin.		
XX			
KW	Human; cancer; hepsin; hepsin protease; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200162271-A1.		
XX			
PD	30-AUG-2001.		
XX			
PF	20-FEB-2001; 2001WO-US005703.		
XX			
PR	22-FEB-2000; 2000US-00510738.		
XX			
PA	(UVAR-) UNIV ARKANSAS.		
XX			
FI	O'Brien TJ;		
XX			
DR	WPI; 2001-582004/65.		
XX			
PT	New oligonucleotide complementary to hepsin encoding sequence, useful for		
XX	treating cancer and screening for compounds that inhibit hepsin.		
PS	Claim 35; Page 138-139; 126p; English.		
XX			

The invention relates to an oligonucleotide (i) having a complementary sequence to hepsin. The oligonucleotide is useful for the detection of cancer, treatment of cancer and screening for compounds that inhibit hepsin activity. Hepsin protease, mRNA and immunospecific anti-hepsin antibodies are useful for the diagnosis of cancer in an individual.

CC AAS43076-AAS43103 represent coding sequences and oligonucleotides used in the methods of the invention

CC

XX Sequence 1783 BP; 304 A; 580 C; 558 G; 341 T; 0 U; 0 Other;

Query Match 67.4%; Score 1593.6; DB 5; Length 1783;

Best Local Similarity 99.1%; Pred. No. 7.2e-313;

Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 736 CCCCCAGCCCTGCTCCCTCCATCTCTCAGAGTCCCAACCTGCGCCAGAGGTCA 795
 DB 157 CTTGAGAGCTCCGCCCCCAGCTGTCGAGCCCA-GGGTCCCACTGCGCCAGAGGTCA 215
 QY 796 GCCAGGAAATCATTAACAAGAGAGCATGACATGCGCGAGAGAGGGTGGCCGAGCTGTG 855
 DB 216 GCCAGGAAATCATTAACAAGAGAGCATGACATGCGCGAGAGAGGGTGGCCGAGCTGTG 275
 QY 856 CCATGCTGCTCAGACCCAGAGTGGCAGCTCTCACTGCGGGACCTGTCTACTTCTGACA 915
 DB 276 CCATGCTGCTCAGACCCAGAGTGGCAGCTCTCACTGCGGGACCTGTCTACTTCTGACA 335
 QY 916 GCCATGCGGGGCGGATCCTGGGCGCATTTGGCTGTCTCTCAGAGTGGACAGAGACCG 975
 DB 336 GCCATGCGGGGCGGATCCTGGGCGCATTTGGCTGTCTCTCAGAGTGGACAGAGACCG 395
 QY 976 CTGTACCAGTGCAGTGCAGCTCTGCGGACGCTCGGCTCATGTGCTTTTGAACAAGCGAA 1035
 DB 396 CTGTACCAGTGCAGTGCAGCTCTGCGGACGCTCGGCTCATGTGCTTTTGAACAAGCGAA 455
 QY 1036 GGGACGTGGCGGCTGTGTGCTCTCGGCTCCAACGCGCAGGTAGCGCGACTCAGCTGC 1095
 DB 456 GGGACGTGGCGGCTGTGTGCTCTCGGCTCCAACGCGCAGGTAGCGCGACTCAGCTGC 515
 QY 1096 GAGAGATGGGCTCTCAAGGCGCATGCCACCTCGAGCTGGACGTGGACAGCGCGGCG 1155
 DB 516 GAGAGATGGGCTCTCAAGGCGCATGCCACCTCGAGCTGGACGTGGACAGCGCGGCG 575
 QY 1156 GCCAATGGCAGTGGGCTTTCTGTGTGAGACGAGGAGGAGCTGCCCAACCCAGAGG 1215
 DB 576 GCCAATGGCAGTGGGCTTTCTGTGTGAGACGAGGAGGAGCTGCCCAACCCAGAGG 635
 QY 1216 CTGTGGAAGTCACTCTCGTGTGATTTGCCAGAGCGGTTCTTGGCCGCTATGTC 1275
 DB 636 CTGTGGAAGTCACTCTCGTGTGATTTGCCAGAGCGGTTCTTGGCCGCTATGTC 695
 QY 1276 CAAGACTGTGGCGCGCAGGAAGCTGCCGCTGGACCGCATTCGTGGAGGCGCGGAGACAGC 1335
 DB 696 CAAGACTGTGGCGCGCAGGAAGCTGCCGCTGGACCGCATTCGTGGAGGCGCGGAGACAGC 755
 QY 1336 TTGGGCGCGTGGCGCGTGGCAGTCAAGCTTCGATATGAGAGACACTCTGTGGGGA 1395
 DB 756 TTGGGCGCGTGGCGCGTGGCAGTCAAGCTTCGATATGAGAGACACTCTGTGGGGA 815
 QY 1396 TCCCTGCTCTCCGGGAGCTGGGTCTGACAGCGCGCCCATCTTCCGAGACGGACCGG 1455
 DB 816 TCCCTGCTCTCCGGGAGCTGGGTCTGACAGCGCGCCCATCTTCCGAGACGGACCGG 875
 QY 1456 GTCCGTGCCGATGGCAGATTTTGGCGGTGGCGGTGGCCAGAGGCTCTCCCAAGGCTG 1515
 DB 876 GTCCGTGCCGATGGCAGATTTTGGCGGTGGCGGTGGCCAGAGGCTCTCCCAAGGCTG 935
 QY 1516 CAGCTGGGGGTGACAGCTGTGTCTACACAGGGGCTATCTTCCCTTTCGGAACCCCAAC 1575
 DB 936 CAGCTGGGGGTGACAGCTGTGTCTACACAGGGGCTATCTTCCCTTTCGGAACCCCAAC 995
 QY 1576 AGCGAGAGAAACAGCAACGATATTCCTGTGTCACCTTTCAGTCCCTGCGCCCTACA 1635
 DB 996 AGCGAGAGAAACAGCAACGATATTCCTGTGTCACCTTTCAGTCCCTGCGCCCTACA 1055

QY 1636 GAATACATCCAGCTGTGTGCTCTCCAGCTGCGCGCCAGGCGCTGTGTGATGGCAAGATC 1695
 DB 1056 GAATACATCCAGCTGTGTGCTCTCCAGCTGCGCGCCAGGCGCTGTGTGATGGCAAGATC 1115
 QY 1696 TGTACCTGTGACGGGCTGGGCAACACGACAGTACTATGGCCCAAGCGCGGGTATCTCAG 1755
 DB 1116 TGTACCTGTGACGGGCTGGGCAACACGACAGTACTATGGCCCAAGCGCGGGTATCTCAG 1175
 QY 1756 GAGGCTGTGATCCCATATACGATGATGTGTGCAATGGCGCTGACTTCTATGGAAC 1815
 DB 1176 GAGGCTGTGATCCCATATACGATGATGTGTGCAATGGCGCTGACTTCTATGGAAC 1235
 QY 1816 CAGATCAAGCCCAAGATGTTCTGTGCTGAGTACCCCGAGGGGTGGCATTTGCTGCGAG 1875
 DB 1236 CAGATCAAGCCCAAGATGTTCTGTGCTGAGTACCCCGAGGGGTGGCATTTGCTGCGAG 1295
 QY 1876 GGGACACAGCGGTGTCCTTTGTGTGTGAGAGACAGATCTCTGAGCGCCAGCTTGGCGG 1935
 DB 1296 GGGACACAGCGGTGTCCTTTGTGTGTGAGAGACAGATCTCTGAGCGCCAGCTTGGCGG 1355
 QY 1936 CTGTGTGGCATTTGTGATTTGGGCACTGGCTGTGCTTGGCCCGAAGCCAGCGCTTAC 1995
 DB 1356 CTGTGTGGCATTTGTGATTTGGGCACTGGCTGTGCTTGGCCCGAAGCCAGCGCTTAC 1415
 QY 1996 ACCAAGTCAAGTACTTCCGGAGTGGATCTTCAGAGCCATAAAGACTCACTCCGAAGCC 2055
 DB 1416 ACCAAGTCAAGTACTTCCGGAGTGGATCTTCAGAGCCATAAAGACTCACTCCGAAGCC 1475
 QY 2056 AGCGGCAATGGTGAACCAAGCTCTGACCGGTGGCTTCTGCTGCGCAGCTTCAGGGCCGA 2115
 DB 1476 AGCGGCAATGGTGAACCAAGCTCTGACCGGTGGCTTCTGCTGCGCAGCTTCAGGGCCGA 1535
 QY 2116 GGTGATCCCGTGTGGAGTTCACAGCTGGGCGAGATGGAGATTTTCTTCTTGGGCC 2175
 DB 1536 GGTGATCCCGTGTGGAGTTCACAGCTGGGCGAGATGGAGATTTTCTTCTTGGGCC 1595
 QY 2176 CGGTCAAGTGCACCAAGACACCTCCCTCAGAGGCTCTCTTCCACAGTGGCGGCGCC 2235
 DB 1596 CGGTCAAGTGCACCAAGACACCTCCCTCAGAGGCTCTCTTCCACAGTGGCGGCGCC 1655
 QY 2236 ACTCAGCCCGGAGACCAACCACTCACTCTGACCCCGCATATATATTTGCTGCT 2295
 DB 1656 ACTCAGCCCGGAGACCAACCACTCACTCTGACCCCGCATATATATTTGCTGCT 1715
 QY 2296 GTCTGGAGCTCTGTCTAGGTGCCCTGATGATGGAGTCTTTTAATTAATAAGATGG 2355
 DB 1716 GTCTGGAGCTCTGTCTAGGTGCCCTGATGATGGAGTCTTTTAATTAATAAGATGG 1775
 QY 2356 TTTTGATT 2363
 DB 1776 TTTTGATT 1783

RESULT 12

AB079532 ID AB079532 standard; cDNA; 1783 BP.

XX AB079532;

XX 25-NOV-2002 (first entry)

XX Human hepsin polypeptide encoding cDNA.

XX Hepsin; cancer; tumour; cytosolic; gene therapy; human; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX FT CDS 246..1499

XX FT /product= "hepsin"

PN WO200264839-A2.
XX
XX 22-AUG-2002.
XX
XX 12-FEB-2002; 2002WO-US004018.
XX
XX 14-FEB-2001; 2001US-0268361P.
XX
XX (TULSA-) TULARK INC.
XX
XX Mu D, Powers S;
XX
XX WPI; 2002-643496/69.
XX
XX P-P8DB; ABB01197.
XX
XX
XX Diagnosing a cancer in a mammal, useful for preventing and treating
XX cancer, e.g. ovarian, prostate, breast or lung cancer, comprises
XX detecting and measuring the hepsin gene copy number and comparing the
XX data obtained to a control data.
XX
XX
XX Claim 25; Page 77; 77pp; English.
XX
XX The invention relates to diagnosing cancer in a mammal by detecting the
XX hepsin gene. The method involves detecting and measuring the hepsin gene
XX copy number in a biological subject from a region of the mammal that is
XX suspected to be precancerous or cancerous to generate data for a test
XX gene copy, and comparing the test gene copy number to a data for a
XX control gene copy number, where an amplification of the gene in the
XX biological subject relative to the control indicates the presence of a
XX precancerous lesion or cancer in the animal. The method is useful for
XX diagnosing, preventing and treating cancer, e.g. ovarian, prostate,
XX breast or lung cancer, by the detection of the hepsin gene. The hepsin
XX genes and their expressed protein products are useful as targets for
XX therapy and for identifying compounds useful in the diagnosis, prevention
XX and treatment of tumours and cancers. The present sequence represents the
XX human hepsin polypeptide encoding cDNA (GenBank Accession No. M18930)
XX
XX
XX Sequence 1783 BP; 304 A; 580 C; 558 G; 341 T; 0 U; 0 Other;
XX
XX
XX Query Match 67.4%; Score 1593.6; DB 6; Length 1783;
XX Best Local Similarity 99.1%; Pred. No. 7.2e-313;
XX Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
XX
XX
XX 736 CCCCCAGGCTCTCCCTCCATCTCTCTCAAGGTCCCACTGCGCCAGAGAGTCA 795
XX 157 CTTGAGAGCTCCGCCCACTGCTGACCCCA-GGGTCCCACTGCGCCAGAGGTCA 215
XX 796 GCCAGGAGATCATTAACAAGAGGAGTGAATGGGCGAAGAGAGGGGCGGACTGTG 855
XX 216 GCCAGGAGATCATTAACAAGAGGAGTGAATGGGCGAAGAGAGGGGCGGACTGTG 275
XX 856 CCATGCTGCTCAGACCCAGAGGTGAGCTCTCACTGCGGAGACCTGCTACTTCTGACA 915
XX 276 CCATGCTGCTCAGACCCAGAGGTGAGCTCTCACTGCGGAGACCTGCTACTTCTGACA 335
XX 916 GCCATCGGGGGGAGCATCTGCGGCAATGTGCTGTTCTCTCAGAGAGTGAACAGAGCCG 975
XX 336 GCCATCGGGGGGAGCATCTGCGGCAATGTGCTGTTCTCTCAGAGAGTGAACAGAGCCG 395
XX 976 CTGATACCAAGTGAAGTCAAGTCTGCGGAGCTGCGGCTCATGATCTTTGACAAAGGAA 1035
XX 396 CTGATACCAAGTGAAGTCAAGTCTGCGGAGCTGCGGCTCATGATCTTTGACAAAGGAA 455
XX 1036 GGAAGCTGAGGAGCTGCTGTGCTCTCTCGGCTCAACGCGCAGAGGTAGCGGACTGCTG 1095
XX 456 GGAAGCTGAGGAGCTGCTGTGCTCTCTCGGCTCAACGCGCAGAGGTAGCGGACTGCTG 515
XX 1096 GAGAGATGAGCTTCTCTAGGCACTGACCACTTCCAGCTGAGAGTGAACGCGGCGG 1155
XX 516 GAGAGATGAGCTTCTCTAGGCACTGACCACTTCCAGCTGAGAGTGAACGCGGCGG 575
XX 1156 GCCAATGAGCAGTGGGCTTCTCTGATGAGAGGAGGAGGAGGCGCCCAACCCAGAGG 1215

Db 576 GCCAATGAGCAGTGGGCTTCTCTGATGAGAGGAGGAGGCTGCCCAACCCAGAGG 635
Qy 1216 CTGCTGAGAGTCACTCTCCGTGTGATTTGCCCAAGAGCGGTTCTTGTGCGCATCTGC 1275
Db 636 CTGCTGAGAGTCACTCTCCGTGTGATTTGCCCAAGAGCGGTTCTTGTGCGCATCTGC 695
Qy 1276 CAAGACTGTGAGCGCAGAGAGCTGCGCGTGAACCGCATCGGAGGAGCGGAGACACGAGC 1335
Db 696 CAAGACTGTGAGCGCAGAGAGCTGCGCGTGAACCGCATCGGAGGAGCGGAGACACGAGC 755
Qy 1336 TTGGGCGGAGTGGCGGTGAGTCAAGTCAAGCTTCTGATGATGAGAGACACTCTGTGGGGA 1395
Db 756 TTGGGCGGAGTGGCGGTGAGTCAAGTCAAGCTTCTGATGATGAGAGACACTCTGTGGGGA 815
Qy 1396 TCCCTGCTCTCCGGGAGCTGGGTCTGACAGCGGCCCACTGCTTCCGGAGCGGAACTGG 1455
Db 816 TCCCTGCTCTCCGGGAGCTGGGTCTGACAGCGGCCCACTGCTTCCGGAGCGGAACTGG 875
Qy 1456 GTCCGTCCCGAGTGGAGTGTGGCGGAGTGGCGGAGCGGAGCGGAGCGGAGCGGAGCTG 1515
Db 876 GTCCGTCCCGAGTGGAGTGTGGCGGAGTGGCGGAGCGGAGCGGAGCGGAGCGGAGCTG 935
Qy 1516 CAGCTGGGAGTGAAGGCTGTGCTTACACAGGAGGCTATCTTCCCTTTCGGGAGCCCAAC 1575
Db 936 CAGCTGGGAGTGAAGGCTGTGCTTACACAGGAGGCTATCTTCCCTTTCGGGAGCCCAAC 995
Qy 1576 AGCAGAGAGAAACAGAAACATATTGCTTGTGCTTCACTTCCAGTCCCTGCGCTTCA 1635
Db 996 AGCAGAGAGAAACAGAAACATATTGCTTGTGCTTCACTTCCAGTCCCTGCGCTTCA 1055
Qy 1636 GAATACATCAAGCTGTGTGCTTCCAGCTGCGGAGCGGAGCGGAGTGGAGTGGCAATC 1695
Db 1056 GAATACATCAAGCTGTGTGCTTCCAGCTGCGGAGCGGAGCGGAGTGGAGTGGCAATC 1115
Qy 1696 TGTACCTGAGCGGCTGAGGAGCAACAGCAGTATGATGAGCAAGCGGAGTACTCAG 1755
Db 1116 TGTACCTGAGCGGCTGAGGAGCAACAGCAGTATGATGAGCAAGCGGAGTACTCAG 1175
Qy 1756 GAGGCTGAGTCCCATATACAGATGATGTGCAATGGCGCTGACTTATGAAAC 1815
Db 1176 GAGGCTGAGTCCCATATACAGATGATGTGCAATGGCGCTGACTTATGAAAC 1235
Qy 1816 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGGAGTGGATGATGCTGCTGAG 1875
Db 1236 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGGAGTGGATGATGCTGCTGAG 1295
Qy 1876 GAGCAGACGAGTGGCTTCTTGTGTGAGGAGCAGCATCTCTCGAGCGCCAGCTTGGCG 1935
Db 1296 GAGCAGACGAGTGGCTTCTTGTGTGAGGAGCAGCATCTCTCGAGCGCCAGCTTGGCG 1355
Qy 1936 CTGTGTGAGTGTGATGAGTGGGAGCTGAGCTGTGCTGCGCCAGAGCGGCTTAC 1995
Db 1356 CTGTGTGAGTGTGATGAGTGGGAGCTGAGCTGTGCTGCGCCAGAGCGGCTTAC 1415
Qy 1996 ACCAAATCAGTACTTCCGAGAGTGAATCTTCAAGGCGCATTAAGACTCACTCGAAGCC 2055
Db 1416 ACCAAATCAGTACTTCCGAGAGTGAATCTTCAAGGCGCATTAAGACTCACTCGAAGCC 1475
Qy 2056 AGCGGAGTGTGAGCAGCTCTGACCGGAGTGTCTTCTGCTGCGAGCTCAAGGAGCCGA 2115
Db 1476 AGCGGAGTGTGAGCAGCTCTGACCGGAGTGTCTTCTGCTGCGAGCTCAAGGAGCCGA 1535
Qy 2116 GGTGATCCCGGTGTGGATTCACGCTGGGCGAGAGTGGAGCTTTTCTTCTTGGGCG 2175
Db 1536 GGTGATCCCGGTGTGGATTCACGCTGGGCGAGAGTGGAGCTTTTCTTCTTGGGCG 1595
Qy 2176 CGGTTCACAGTTCACAGAGAACCTCTCCCTCAGAGTCTCTTCTTCACTATGGGCGG 2235
Db 1596 CGGTTCACAGTTCACAGAGAACCTCTCCCTCAGAGTCTCTTCTTCACTATGGGCGG 1655
Qy 2236 ACTCAGGCGGAGACCAACCACTCTCAGTCCGAGTCAAGTAAATATGTTCTGCT 2295
Db 1656 ACTCAGGCGGAGACCAACCACTCTCAGTCCGAGTCAAGTAAATATGTTCTGCT 1715


```
Db 1236 CAGATCAAGCCCAAGATTTCTGTGCTGTACACCCGAGGGTGGATTGATGCTCCAG 1235
Qy 1876 GCGCACAAGCGGTGTCCTTTGTGTGTAGAGACAGCATCTCTCGAGCCACGTTGGCG 1935
Db 1296 GCGCACAAGCGGTGTCCTTTGTGTGTAGAGACAGCATCTCTCGAGCCACGTTGGCG 1355
Qy 1336 CTGTGTGCGATTGTGAGTTGGGACCTGGCTGTGCTGCTGCGCCAGAAAGCGGCTTAC 1935
Db 1356 CTGTGTGCGATTGTGAGTTGGGACCTGGCTGTGCTGCTGCGCCAGAAAGCGGCTTAC 1415
Qy 1596 ACCAAGCTAGTGACTTCCGGAGTGTATCTTCCAGGCGCATTAAGACTCACTCCGAAAGCC 2055
Db 1416 ACCAAGCTAGTGACTTCCGGAGTGTATCTTCCAGGCGCATTAAGACTCACTCCGAAAGCC 1475
Qy 2056 AGCGGCGATGTGACCCAGCTGTGACCGGTGCTTCTGCTGCGACGCTCCAGGCGCGCA 2115
Db 1476 AGCGGCGATGTGACCCAGCTGTGACCGGTGCTTCTGCTGCGACGCTCCAGGCGCGCA 1535
Qy 2116 GGTGATCCCGGTGTGGGATTCACGCTGTGGGCGGAGATGGACGTTTTCTTCTTGGGCC 2175
Db 1536 GGTGATCCCGGTGTGGGATTCACGCTGTGGGCGGAGATGGACGTTTTCTTCTTGGGCC 1595
Qy 2176 CGGTCCACAGGTCCAGACACACCTCCCTCCAGGCTCTCTTCCACAGTGGCGGCGCC 2235
Db 1596 CGGTCCACAGGTCCAGACACACCTCCCTCCAGGCTCTCTTCCACAGTGGCGGCGCC 1655
Qy 2236 ACTGAGCGCCGAGACCAACCACTCACTCTGACCCCGATGTAATATTTGTTGCT 2295
Db 1656 ACTGAGCGCCGAGACCAACCACTCACTCTGACCCCGATGTAATATTTGTTGCT 1715
Qy 2296 GTCTGGGACTCTGTCTTAGTGTGCTGCTGTATGATGGAGTCTTTTAATATAAAGATGG 2355
Db 1716 GTCTGGGACTCTGTCTTAGTGTGCTGCTGTATGATGGAGTCTTTTAATATAAAGATGG 1775
Qy 2356 TTTTGATT 2363
Db 1776 TTTTGATT 1783

RESULT 14
ADB75353
ID ADB75353 standard; cDNA; 1783 BP.
XX
AC ADB75353;
XX
DT 04-DEC-2003 (first entry)
XX
DE Prostate cancer marker cDNA.
XX
KW Prostate; cancer; cytostatic; gene therapy; marker; ss.
XX
OS Homo sapiens.
XX
PN WO2003009814-A2.
XX
PD 06-FEB-2003.
XX
PF 25-JUL-2002; 2002WO-US023913.
XX
PR 25-JUL-2001; 2001US-0307982P.
PR 22-AUG-2001; 2001US-0314356P.
PR 25-SEP-2001; 2001US-0325020P.
PR 12-DEC-2001; 2001US-0341746P.
PR 05-MAR-2002; 2002US-0362158P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
PI Schlegel R, Monahan JE, Endege WO, Gannavarapu M, Gorbacheva B,
XX Hoeroh S, Kamatkar S, Womsey AM, Glatt K, Zhao X, Anderson D,
XX WPI; 2003-248033/24.
XX
```

```
PT New nucleic acid molecule, useful for diagnosing or treating prostate
PT cancer.
XX
PS Disclosure; SEQ ID NO 177; 99pp; English.
XX
XX The invention relates to newly discovered cancer markers associated with
CC the cancerous state of prostate cells. Also disclosed is a method of
CC assessing whether a patient is afflicted with prostate cancer. The method
CC of the invention involves assessing whether a patient is afflicted with
CC prostate cancer by comparing the level of expression of a marker in a
CC patient sample and the normal level of expression of the marker in a
CC control non-prostate cancer sample, where a significant increase in the
CC level of expression of the marker in the patient sample and the normal
CC level indicates that the patient is afflicted with prostate cancer.
CC Nucleic acids of the invention are useful for diagnosing or treating
CC prostate cancer, and may be useful in gene therapy. Sequences given in
CC ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1783 BP; 304 A; 580 C; 558 G; 341 T; 0 U; 0 Other;
SQ
Query Match 67.4%; Score 1593.6; DB 10; Length 1783;
Best Local Similarity 99.1%; Pred. No. 7.2e-313;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
Qy 736 CCCCCAGGCCCTGCTCCGCTCCATCTCTCAAGTCCACCTGCGCCAGAGGTCA 795
Db 157 CTTGAGAGCTCCGCCCCCAGCTGTGAGACCCA-GGGTCCACCTGCGCCAGAGGTCA 215
Qy 796 GCCAGGGAATCATTAACAAGAGCAATGACATGGCCGAGAAAGGTGGCCGACCTGTG 855
Db 216 GCCAGGGAATCATTAACAAGAGCAATGACATGGCCGAGAAAGGTGGCCGACCTGTG 275
Qy 856 CCATGCTGCTCCAGACCCAGAGTGGAGCTCTCACTCGGAGGACCTGTCTACTTGTACA 915
Db 276 CCATGCTGCTCCAGACCCAGAGTGGAGCTCTCACTCGGAGGACCTGTCTACTTGTACA 335
Qy 916 GCCATCGGGGCGGATCTGGGCCATTTGGGCTGTTCTCTCAGAGTGAACAGAGCCG 975
Db 336 GCCATCGGGGCGGATCTGGGCCATTTGGGCTGTTCTCTCAGAGTGAACAGAGCCG 395
Qy 976 CTGTACCAAGTGAAGTCTCACTCTGCGGAGAGCTGTGCTCATGTCTTTGACAAAGCGGAA 1035
Db 396 CTGTACCAAGTGAAGTCTCACTCTGCGGAGAGCTGTGCTCATGTCTTTGACAAAGCGGAA 455
Qy 1036 GGGACGTGGGCGGCTGTGTCTCTCGCGCTCCAGCCGAGGAGGCGGAGCTCAGCTGC 1095
Db 456 GGGACGTGGGCGGCTGTGTCTCTCGCGCTCCAGCCGAGGAGGCGGAGCTCAGCTGC 515
Qy 1096 GAGAGAGTGGGCTTCTCTCAGAGGCACTGACCCACTCCAGTGTGAAGTGTGAGCGGCGGC 1155
Db 516 GAGAGAGTGGGCTTCTCTCAGAGGCACTGACCCACTCCAGTGTGAAGTGTGAGCGGCGGC 575
Qy 1156 GCCAATGGCACTGTGGGCTTCTTGTGTGTGAGAGAGGAGGCTGCCCAACCCAGAGG 1215
Db 576 GCCAATGGCACTGTGGGCTTCTTGTGTGTGAGAGAGGAGGCTGCCCAACCCAGAGG 635
Qy 1216 CTGTGAGAGTCACTCTCGGTGTGATTTGCCCCAGAGGCGGTTTCTTGGCGGCACTGCG 1275
Db 636 CTGTGAGAGTCACTCTCGGTGTGATTTGCCCCAGAGGCGGTTTCTTGGCGGCACTGCG 695
Qy 1276 CAAGACTGTGGCCGAGAGAGTGTCCCGTGAACCGCATCTGTGGAGGCGCGGAGACACGAGC 1335
Db 696 CAAGACTGTGGCCGAGAGAGTGTCCCGTGAACCGCATCTGTGGAGGCGCGGAGACACGAGC 755
Qy 1336 TTGGGCGGCTGGCGCGGAGTCAAGCTCTTCCGTATGATGAGACACCTCTGTGGGGGA 1395
Db 756 TTGGGCGGCTGGCGCGGAGTCAAGCTCTTCCGTATGATGAGACACCTCTGTGGGGGA 815
Qy 1396 TCCCTGCTCCGAGGAGCTGGGTGTGACAGCGGCGCACTGTCTTCCGAGAGCGGAAACGG 1455
Db 815
```

```
Db      816  TCCCTGCTCTCCGGGAGACTGGGTGTGACAGCCGCCCACTGCTTCCGGAGCGGAACCGG  875
Qy      1456 GTCTGTCCCGATGGCCAGATGTTTGCCGGTCCCTGGGCCCAAGGCTCTCCCCAAGGTCTG  1515
Db      876  GTCTGTCCCGATGGCCAGATGTTTGCCGGTCCCTGGGCCCAAGGCTCTCCCCAAGGTCTG  935
Qy      1516 CAGCTGGGGGGGAGAGCTGTGGTCTACCAACGGGGGGCTATCTTCCCTTCCGGAGCCCAAC  1575
Db      936  CAGCTGGGGGGGAGAGCTGTGGTCTACCAACGGGGGGCTATCTTCCCTTCCGGAGCCCAAC  995
Qy      1576 AGCGAGGAGAAACAGCAACGATATTTGCCCTGGTCCACTCTCAGTCCCTGCCCCCTACA  1635
Db      996  AGCGAGGAGAAACAGCAACGATATTTGCCCTGGTCCACTCTCAGTCCCTGCCCCCTACA  1055
Qy      1636 GAATACATCCAGCTGTGTGCTCTCCCACTGCCGCCAGGGCCCTGGTGGATGGCAAGATC  1695
Db      1056 GAATACATCCAGCTGTGTGCTCTCCCACTGCCGCCAGGGCCCTGGTGGATGGCAAGATC  1115
Qy      1696 TGTACCGTGAAGGGCTGGGGGCAACAGCAGTACTATGGCCAAACAGGCCGGGGGTACTCCAG  1755
Db      1116 TGTACCGTGAAGGGCTGGGGGCAACAGCAGTACTATGGCCAAACAGGCCGGGGGTACTCCAG  1175
Qy      1756 GAGGCTCGAGTCCCATATATCAGCAATGATGTGCAATGGCGCTGACTTCTATGGAAC  1815
Db      1176 GAGGCTCGAGTCCCATATATCAGCAATGATGTGCAATGGCGCTGACTTCTATGGAAC  1235
Qy      1816 CAGATCAAGCCCAAGATGTTGTGTGCTGGTACCCCGAGGGTGGCATTTGATGCTGCGAC  1875
Db      1236 CAGATCAAGCCCAAGATGTTGTGTGCTGGTACCCCGAGGGTGGCATTTGATGCTGCGAC  1295
Qy      1876 GGCACAGACGGGTGTCTCCCTTGTGTGAGAGACAGCATCTCTCGAAGCGCCAGTTGGCGG  1935
Db      1296 GGCACAGACGGGTGTCTCCCTTGTGTGAGAGACAGCATCTCTCGAAGCGCCAGTTGGCGG  1355
Qy      1936 CTGTGTGCGATTTGAGTTGGGGGCACTGGCTGTGCTGCCCTCGCCAGAACCGAGCGTCTAC  1995
Db      1356 CTGTGTGCGATTTGAGTTGGGGGCACTGGCTGTGCTGCCCTCGCCAGAACCGAGCGTCTAC  1415
Qy      1996 ACCAAAGTCAGTGACTTCCGGGAGTGGATCTTCCAGGCGCAATGAAGACTCACTCCGAAGCC  2055
Db      1416 ACCAAAGTCAGTGACTTCCGGGAGTGGATCTTCCAGGCGCAATGAAGACTCACTCCGAAGCC  1475
Qy      2056 AGCGGCAATGTGACCCAGCTGTGACCGGTGGCTTCTCGCTGCGAGCTCCAGGGGCCGA  2115
Db      1476 AGCGGCAATGTGACCCAGCTGTGACCGGTGGCTTCTCGCTGCGAGCTCCAGGGGCCGA  1535
Qy      2116 GTGTATCCCGGTGTGGGATCAGCGTGGGCGAGGATGGAGCGTTTTTCTTCTTGGGCC  2175
Db      1536 GTGTATCCCGGTGTGGGATCAGCGTGGGCGAGGATGGAGCGTTTTTCTTCTTGGGCC  1595
Qy      2176 CGGTCCACAGTCCCAAGGACACCTCCCTCCAGGGTCTCTCTTCCACAGTGGGGGCC  2235
Db      1596 CGGTCCACAGTCCCAAGGACACCTCCCTCCAGGGTCTCTCTTCCACAGTGGGGGCC  1655
Qy      2236 ACTCAGCCCCGAGACCAACCACTCAACCTCTGACCCGCCCATATTAATATTTGCTGCT  2295
Db      1656 ACTCAGCCCCGAGACCAACCACTCAACCTCTGACCCGCCCATATTAATATTTGCTGCT  1715
Qy      2296 GTCTGGGACTCTGTCTAGAGTGGCCCTGGATGGAGATGCTCTTTAAATAAATAAAGATGG  2355
Db      1716 GTCTGGGACTCTGTCTAGAGTGGCCCTGGATGGAGATGCTCTTTAAATAAATAAAGATGG  1775
Qy      2356 TTTTGATT 2363
Db      1776 TTTTGATT 1783
```

```
RESULT 15
ADD18429
ID      ADD18429 standard; DNA; 1783 BP.
XX
XX      ADD18429;
XX
```

```
DT      15-JAN-2004 (first entry)
XX
DE      Human prostate cancer diagnosis related DNA sequence SegID1.
XX
KW      prostate tissue; cancer diagnostic; cancer marker; prostate cancer; PCA;
KW      male cancer-related death; serum biomarker; tissue biomarker; cytostatic;
KW      gene therapy; prostate biopsy tissue; AMACR;
KW      alpha-methylacyl-coenzyme A racemase; diagnosing cancer; cell growth;
KW      human; de.
XX
OS      Homo sapiens.
XX
PN      WO2003012067-A2.
XX
PD      13-FEB-2003.
XX
PF      02-AUG-2002; 2002WO-US024567.
XX
PR      02-AUG-2001; 2001US-0309581P.
PR      15-NOV-2001; 2001US-0334468P.
PR      01-AUG-2002; 2002US-00210120.
XX
PA      (UNMI ) UNIV MICHIGAN.
XX
PI      Rubin MA, Chinaaiyan AM, Sreekumar A;
XX
DR      WPI; 2003-278396/27.
XX
PT      Characterizing prostate tissue comprises providing a prostate tissue
PT      sample from a subject and detecting the presence or absence of expression
PT      of hepsin, pim-1 or Ezh2.
XX
PS      Disclosure; SEQ ID NO 1; 297bp; English.
XX
CC      This invention relates to a novel method of characterizing prostate
CC      tissue in a subject and to compositions and methods for cancer
CC      diagnostics, including cancer markers, in particular prostate cancer.
CC      Prostate cancer (PCA) is a leading cause of male cancer-related death.
CC      Additional serum and tissue biomarkers would aid diagnosis. The invention
CC      may provide means of producing compounds with a cytostatic activity or
CC      allow the development of gene therapy. The methods of the invention
CC      useful for characterizing prostate tissue in a subject, screening
CC      compounds, characterizing inconclusive prostate biopsy tissue in a
CC      subject, detecting AMACR (alpha-methylacyl-coenzyme A racemase)
CC      expression in a bodily fluid, characterizing tissue in a subject,
CC      diagnosing cancer in a subject and inhibiting the growth of cells. The
CC      present sequence is a DNA sequence which is preferably utilised in the
CC      method of the invention.
XX
SQ      Sequence 1783 BP; 304 A; 580 C; 558 G; 341 T; 0 U; 0 Other;
XX
Query Match 67.4%; Score 1593.6; DB 10; Length 1783;
Best Local Similarly 99.1%; Pred. No. 7.2e-313;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
Qy      736 CCCCAGAGCCCTGCTCCCGTCATCTCTCAAGGATCCCACTGGCCAGAGGTCA 795
Db      157 CTTGAGGCTTCGCCCCCACTGTGACCCCA-GGGTCCACCTTGCCCAAGAGTCA 215
Qy      796 GCCAGGGAAATCATTAACAGAGGAGTGAATGAGCGCAGAAAGAGGGTGGCCGACTGTG 855
Db      157 CTTGAGGCTTCGCCCCCACTGTGACCCCA-GGGTCCACCTTGCCCAAGAGTCA 215
Qy      856 CCATGCTGCTCCAGACCCAGAGGTGAGCTCTCACTGCGGGGACCCCTGATCTTGTACA 915
Db      216 GCCAGGGAAATCATTAACAGAGGAGTGAATGAGCGCAGAAAGAGGGTGGCCGACTGTG 275
Qy      916 GGCATGGGGGCGGATCTGGGCAATGTGGCTGTTCTCTCAAGAGTGAACAGAGCCG 975
Db      276 CCAATGCTCTCCAGACCCAGAGGTGAGCTCTCACTGCGGGGACCCCTGATCTTGTACA 335
Qy      976 GGCATGGGGGCGGATCTGGGCAATGTGGCTGTTCTCTCAAGAGTGAACAGAGCCG 975
Db      336 GGCATGGGGGCGGATCTGGGCAATGTGGCTGTTCTCTCAAGAGTGAACAGAGCCG 395
Qy      976 CTGTACCCAGTGAAGTCACTCTGCGAGCGCTCGGCTCATGTGCTTTTGAACAAGCGAA 1035
XX
```

Db 396 CTGTACCCAGTGCAGCTCTGCGGACGCTCGGCTCATGTCTTTGACAAAGACGAA 455
 Qy 1036 GGGAGTGTGGCGGCTGTGTCTCTCGGCTTCCAAACGCCAGGGTATGCCGACTCAGCTGC 1095
 Db 456 GGGAGTGTGGCGGCTGTGTCTCTCGGCTTCCAAACGCCAGGGTATGCCGACTCAGCTGC 515
 Qy 1096 GAGGAGATGGGCTTCTCTCAGGGCACTGACCCCACTCGAGCTGGAAGTGCAGACGGGAGC 1155
 Db 516 GAGGAGATGGGCTTCTCTCAGGGCACTGACCCCACTCGAGCTGGAAGTGCAGACGGGAGC 575
 Qy 1156 GCCAATGGCAGCTCGGGCTTCTCTGTGTGACGAGGGGAGGCTGCCCAACCCAGAGG 1215
 Db 576 GCCAATGGCAGCTCGGGCTTCTCTGTGTGACGAGGGGAGGCTGCCCAACCCAGAGG 635
 Qy 1216 CTGTGTGAGGTTCATCTCGGTGTGTATTTGCCCAAGGGCCGTTTCTTGGCCGCACTTGC 1275
 Db 636 CTGTGTGAGGTTCATCTCGGTGTGTATTTGCCCAAGGGCCGTTTCTTGGCCGCACTTGC 695
 Qy 1276 CAAGACTGTGGCCGCGAGGAGCTGCCGTTGGACCGCATCGTGGAGGCGCGGAGACACAGC 1335
 Db 696 CAAGACTGTGGCCGCGAGGAGCTGCCGTTGGACCGCATCGTGGAGGCGCGGAGACACAGC 755
 Qy 1336 TTGGGCCGCGTGCCTGTGCAAGTCAAGCTTGCCTATGTAGTGAACACACTCTGTGTGGGGA 1395
 Db 756 TTGGGCCGCGTGCCTGTGCAAGTCAAGCTTGCCTATGTAGTGAACACACTCTGTGTGGGGA 815
 Qy 1396 TTCCTGCTCTCCGGGGGACTGGGTGTGACAGCCGCCCACTGCTTCCGGAAGCGAACCCGG 1455
 Db 816 TTCCTGCTCTCCGGGGGACTGGGTGTGACAGCCGCCCACTGCTTCCGGAAGCGAACCCGG 875
 Qy 1456 GTCCGTGTCCGAGTGGCGAGTGTGTTGGCGGGTGCCTGGGCGCAAGGCTCTCCCAAGGTCTG 1515
 Db 876 GTCCGTGTCCGAGTGGCGAGTGTGTTGGCGGGTGCCTGGGCGCAAGGCTCTCCCAAGGTCTG 935
 Qy 1516 CAGCTGGGGGTGTGACAGCTGTGTGTCTTACCAACGCGGGGTATCTTCTTCCGGAACCCCAAC 1575
 Db 936 CAGCTGGGGGTGTGACAGCTGTGTGTCTTACCAACGCGGGGTATCTTCTTCCGGAACCCCAAC 995
 Qy 1576 AGCGAGAGAAACAGCAAGATTTGCCCTGTGCACTTCTTCCAGTCCCTGCCCCCTCACA 1635
 Db 996 AGCGAGAGAAACAGCAAGATTTGCCCTGTGCACTTCTTCCAGTCCCTGCCCCCTCACA 1055
 Qy 1636 GAATACATCCAGCTGTGTGCTCCAGTGCCTGCCGAGCGGCTGTGGATGGCAAGATC 1695
 Db 1056 GAATACATCCAGCTGTGTGCTCCAGTGCCTGCCGAGCGGCTGTGGATGGCAAGATC 1115
 Qy 1696 TGTACCGTGAAGGGCTGGGGCAACACGAGTACTATGGCAACAGGCGGGGTACTCCAG 1755
 Db 1116 TGTACCGTGAAGGGCTGGGGCAACACGAGTACTATGGCAACAGGCGGGGTACTCCAG 1175
 Qy 1756 GAGGCTGAGTCCCAATATAGCAATGATGTTCGAATGGCGCTGACTTCTATGAAAC 1815
 Db 1176 GAGGCTGAGTCCCAATATAGCAATGATGTTCGAATGGCGCTGACTTCTATGAAAC 1235
 Qy 1816 CAGATCAAGCCCAAGATTTCTGTGCTGTGCTAACCCCGAGGGTGTGATGCTCTGCCAG 1875
 Db 1236 CAGATCAAGCCCAAGATTTCTGTGCTGTGCTAACCCCGAGGGTGTGATGCTCTGCCAG 1295
 Qy 1876 GGCACAGCGGTGTCTCTTGTGTGTGAGACAGCATCTCTCGGACGCCAGTGTGGCGG 1935
 Db 1296 GGCACAGCGGTGTCTCTTGTGTGTGAGACAGCATCTCTCGGACGCCAGTGTGGCGG 1355
 Qy 1936 CTGTGTGAGTGTGTGAGTGTGGGGCACTGGCTGTGCTGGCCCAAGGCCAGGCGTCTAC 1995
 Db 1356 CTGTGTGAGTGTGTGAGTGTGGGGCACTGGCTGTGCTGGCCCAAGGCCAGGCGTCTAC 1415
 Qy 1996 ACCAAAGTCAAGTATTCGCGGAGTGTGATCTTCCAGGGCATTAAGACTCACTCCGAAGCC 2055
 Db 1416 ACCAAAGTCAAGTATTCGCGGAGTGTGATCTTCCAGGGCATTAAGACTCACTCCGAAGCC 1475
 Qy 2056 AGCGGCAATGTGACCCAGCTGTGACCGGGTGGCTTCTCGCTGCGAGCTTCCAGGGGCCGA 2115
 Db 1476 AGCGGCAATGTGACCCAGCTGTGACCGGGTGGCTTCTCGCTGCGAGCTTCCAGGGGCCGA 1535

Qy 2116 GGTGATCCCGGTGTGGATTCACAGTGGGCGGAGATGGGACGTTTTTCTTCTTGGGCC 2175
 Db 1536 GGTGATCCCGGTGTGGATTCACAGTGGGCGGAGATGGGACGTTTTTCTTCTTGGGCC 1595
 Qy 2176 CGGTCCACAGGTCCAAAGGACACCTTCCCTCAGGGTCTCTCTTCCACAGTGGCGGGGCC 2235
 Db 1596 CGGTCCACAGGTCCAAAGGACACCTTCCCTCAGGGTCTCTCTTCCACAGTGGCGGGGCC 1655
 Qy 2236 ACTAGCCCGGAGAACACCAACCTCACTCTTGAACCCCGATGTAATATTGTTCTGCT 2295
 Db 1656 ACTAGCCCGGAGAACACCAACCTCACTCTTGAACCCCGATGTAATATTGTTCTGCT 1715
 Qy 2296 GTCTGGGACTCTGTCTAGTGTGCTGCTGATGATGGAATGCTTTAAATATAAAGATGG 2355
 Db 1716 GTCTGGGACTCTGTCTAGTGTGCTGCTGATGATGGAATGCTTTAAATATAAAGATGG 1775
 Qy 2356 TTTTGATT 2363
 Db 1776 TTTTGATT 1783

Search completed: August 17, 2005, 02:48:42
 Job time : 1922 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2005, 02:14:07 ; Search time 11629 Seconds
(without alignments)
7734.618 Million cell updates/sec

Title: US-10-054-498-1

Perfect score: 2363
Sequence: 1 tcggagccgcgttcacagga.....taataaagatgcttgcattc 2363

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: gb_est1.*
2: gb_est2.*
3: gb_hc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gest1.*
9: gb_gest2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1238	52.4	1238	9	AY408324 Homo sapi
2	1060.8	44.9	1225	9	AY408325 Pan trogl
3	1057.6	44.8	1828	3	CR597177 full-leng
4	939.2	39.7	1814	3	AK002694 Mus muscu
5	932.4	39.5	1238	9	AY408326 Mus muscu
6	905.4	38.3	1212	3	CR592189 full-leng
7	831	35.2	846	1	AL558100
8	814	34.4	916	4	BG754441
9	796.8	33.7	940	4	BG754531
10	794	33.6	965	5	BK406458
11	788.2	33.4	1012	7	CK231686
12	766.4	32.4	881	4	BT768246
13	753	31.9	961	5	BK422288
14	747	31.6	874	5	BK427780
15	726.8	30.8	774	4	BT752774
16	705.6	29.9	887	4	BT760912
17	704.4	29.8	717	4	BT759565
18	702.6	29.7	910	7	CN156776
19	680.6	28.8	829	7	CR765441
20	671.8	28.4	837	5	BK325135
21	671.6	28.4	786	5	BK406457
22	668	28.3	688	5	BM994658
23	668	28.3	931	7	CO775189
24	665.2	28.2	701	7	CV029569

c	25	660	27.9	690	5	BU688738	UI-CF-EC1
	26	659.8	27.9	763	4	BT764518	BT764518
	27	652	27.6	911	5	B0645325	AGENCOURT
	28	640.6	27.1	830	4	BT517769	B0645325
	29	635.8	26.9	649	1	AV656299	AV656299
	30	622.8	26.4	786	4	BT765615	BT765615
	31	618.6	26.2	775	4	BT836887	BT836887
	32	616.6	26.1	698	6	CD614168	CD614168
	33	616.4	26.1	984	4	BT762697	BT762697
	34	612.8	25.9	812	7	CN154671	CN154671
	35	603.8	25.6	1142	4	BM548649	BM548649
	36	603.4	25.5	628	6	CD614167	CD614167
	37	599.8	25.4	609	7	CN272681	CN272681
	38	596	25.2	738	5	BR444180	BR444180
	39	593.6	25.1	930	1	AL528314	AL528314
	40	590.8	25.0	681	2	AM008459	AM008459
	41	570	24.1	879	4	BT764734	BT764734
	42	563.6	23.9	651	1	AL561243	AL561243
	43	560.8	23.7	1116	1	AL557891	AL557891
	44	558.2	23.6	564	1	AL699708	AL699708
	45	552.6	23.4	747	4	BT765513	BT765513

ALIGNMENTS

RESULT 1
AY408324
LOCUS Homo sapiens HPN gene, 1238 bp DNA linear GSS 15-DEC-2003
DEFINITION genomic survey sequence.
ACCESSION AY408324
VERSION AY408324.1 GI:39764295
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1238)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES
source location/Qualifiers
1..1238
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1238
/gene="HPN"
/locus_tag="HCM3186"
ORIGIN
Query Match 52.4%; Score 1238; DB 9; Length 1238;
Best Local Similarity 100.0%; Pred. No. 1.9e-277;
Matches 1238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
842 GTGGCCGACGTCGATGCTGCTCGACACCAAGTGGAGCTCTGACGCGGAGCC 901
|||||

Db 1 GTGACCGGACATGTGCGCATGTCTCAGAACCCAAAGTGGCAGCTCTCAGTCCGGGAGACC 60
 Qy 902 TGTACTTCTTCAAGCCATCGGGGGGGGATCTCTGGGCAATGTGGCTGTCTCTCAGAGA 961
 Db 61 TGTACTTCTTCAAGCCATCGGGGGGGGATCTCTGGGCAATGTGGCTGTCTCTCAGAGA 120
 Qy 962 GTGACCAAGAGACCGCTGTACCCAGTGCAGGTCACTGTGGGAGACGCTCGGCTCAATGCTCT 1021
 Db 121 GTGACCAAGAGACCGCTGTACCCAGTGCAGGTCACTGTGGGAGACGCTCGGCTCAATGCTCT 180
 Qy 1022 TTGAACAAGACGAAAGGAGCGTGGCGGCTGTGTGTCTCTGTGCGCTCCAAAGCCAGGCTAG 1081
 Db 181 TTGAACAAGACGAAAGGAGCGTGGCGGCTGTGTGTCTCTGTGCGCTCCAAAGCCAGGCTAG 240
 Qy 1082 CCGGACTTCAAGTGTGGAAGAGATGGGCTTCTCAGGGGCACTAACCTCCAGCTGAGACG 1141
 Db 241 CCGGACTTCAAGTGTGGAAGAGATGGGCTTCTCAGGGGCACTAACCTCCAGCTGAGACG 300
 Qy 1142 TGGCAACGGGGGCGCCAAATGAGCAAGTGGGCTTCTGTGTGTGAGCAGAGGGAGAGCTGC 1201
 Db 301 TGGCAACGGGGGCGCCAAATGAGCAAGTGGGCTTCTGTGTGTGAGCAGAGGGAGAGCTGC 360
 Qy 1202 CCCAACCCAGAGGCTGTGTGAGGTCACTCTCGTGTGATTTGCCAGAGGCCGTTTCT 1261
 Db 361 CCCAACCCAGAGGCTGTGTGAGGTCACTCTCGTGTGATTTGCCAGAGGCCGTTTCT 420
 Qy 1262 TGGCGCGCATTTGCAAGACTGTGGCCGCAAGAACTGCCGCTGTGACCGGATCGTGGAG 1321
 Db 421 TGGCGCGCATTTGCAAGACTGTGGCCGCAAGAACTGCCGCTGTGACCGGATCGTGGAG 480
 Qy 1322 GCCGGGACACAGGCTTGGGGCCGGTGGCGGTGGCAAGTCAAGCTTCGATGATGAGAGAC 1381
 Db 481 GCCGGGACACAGGCTTGGGGCCGGTGGCGGTGGCAAGTCAAGCTTCGATGATGAGAGAC 540
 Qy 1382 ACTCTGTGGGGAGTCCCTGTCTTCGGGAGACTGGGTGTGACAGACGCGCCACTGTCTCC 1441
 Db 541 ACTCTGTGGGGAGTCCCTGTCTTCGGGAGACTGGGTGTGACAGACGCGCCACTGTCTCC 600
 Qy 1442 CGGAGCGGAAACCGGCTCTGTTCGGGAGTGGCGAGTGGTGGCGGCTGCGCCAGAGGCT 1501
 Db 601 CGGAGCGGAAACCGGCTCTGTTCGGGAGTGGCGAGTGGTGGCGGCTGCGCCAGAGGCT 660
 Qy 1502 CTCCCAACGCTGTGAGCTGGGGGGTGCAGGCTGTGTGATCAACAGGGGGCTATCTCCCT 1561
 Db 661 CTCCCAACGCTGTGAGCTGGGGGGTGCAGGCTGTGTGATCAACAGGGGGCTATCTCCCT 720
 Qy 1562 TTGCGGAGCCCAACAGCGAGGAGAAAGCAACGATATTGCCCTGGTCCACTCTTCCAGTC 1621
 Db 721 TTGCGGAGCCCAACAGCGAGGAGAAAGCAACGATATTGCCCTGGTCCACTCTTCCAGTC 780
 Qy 1622 CCGTGGCCCTCAAGAAATCAATCAAGCTGTGTGCTCCGAGCTGCCGGGAGAGCCCTGG 1681
 Db 781 CCGTGGCCCTCAAGAAATCAATCAAGCTGTGTGCTCCGAGCTGCCGGGAGAGCCCTGG 840
 Qy 1682 TGGATGGCAAGATCTGTACCGTGAAGGCTGGGGCAACAGCCAGTACTATGAGCCAAACAG 1741
 Db 841 TGGATGGCAAGATCTGTACCGTGAAGGCTGGGGCAACAGCCAGTACTATGAGCCAAACAG 900
 Qy 1742 CCGGGGATCTCAAGAGGCTGAGTCCCATATCAAGATGATGTGTGCAATGGCGCTG 1801
 Db 901 CCGGGGATCTCAAGAGGCTGAGTCCCATATCAAGATGATGTGTGCAATGGCGCTG 960
 Qy 1802 ACTCTTATGGAACCAAGTCAAGCCCAAGATGTTCTGTGTGCTTACCCCGAGGGTGGCA 1861
 Db 961 ACTCTTATGGAACCAAGTCAAGCCCAAGATGTTCTGTGTGCTTACCCCGAGGGTGGCA 1020
 Qy 1862 TTGATGCTGTCAGAGGCGACAGCGGTGTCTCTTGTGTGTGAGGACAGCATCTCTCGGA 1921
 Db 1021 TTGATGCTGTCAGAGGCGACAGCGGTGTCTCTTGTGTGTGAGGACAGCATCTCTCGGA 1080
 Qy 1922 CGCCACGTTGGCGGCTGTGTGCAATTTGAGTTGGGGGCACTGCTGTGCTCGGCCAGA 1981
 Db 1081 CGCCACGTTGGCGGCTGTGTGCAATTTGAGTTGGGGGCACTGCTGTGCTCGGCCAGA 1140

Qy 1982 AGCCAGGCGTCTACACCAAGTCAAGTACTTCCGGAGTGGATCTTCCAGGCCATAAGA 2041
 Db 1141 AGCCAGGCGTCTACACCAAGTCAAGTACTTCCGGAGTGGATCTTCCAGGCCATAAGA 1200
 Qy 2042 CTCATCCGGAAGCCAGCGGATGGTGAACCAAGCTCTGA 2079
 Db 1201 CTCATCCGGAAGCCAGCGGATGGTGAACCAAGCTCTGA 1238
 RESULT 2
 AY408325
 LOCUS
 DEFINITION
 genomic survey sequence.
 ACCESSION
 AY408325
 VERSION
 AY408325.1 GI:39764296
 KEYWORDS
 SOURCE
 ORGANISM
 Pan troglodytes (chimpanzee)
 Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
 REFERENCE
 1 (bases 1 to 1225)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
 Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,
 Adams,M.D. and Cargill,M.
 Direct Submision
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 TITLE
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL
 Science 302 (5652), 1960-1963 (2003)
 PUBMED
 14671302
 REFERENCE
 2 (bases 1 to 1225)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
 Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,
 Adams,M.D. and Cargill,M.
 JOURNAL
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT
 This sequence was made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES
 source
 1..1225
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 <1..>1225
 /gene="HPN"
 /locus_tag="HCM3186"
 ORIGIN
 Query Match 44.9%; Score 1060.8; DB 9; Length 1225;
 Best Local Similarity 86.7%; Pred. No. 3,7e-236;
 Matches 1062; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
 Qy 855 GCCATGCTGCTCCAGACCAAGGAGGAGGCTCTCACTGCGGGGACCTGTACTTCTGAC 914
 Db 1 GCATGCTGCTCCAGACCAAGGAGGAGGCTCTCACTGCGGGGACCTGTACTTCTGAC 60
 Qy 915 AGCCATGGGGGCGGCATCTGGGCAATGTGTGCTTCTCTCAGAGTGAACAGAGCC 974
 Db 61 AGCCATGGGGGCGGCATCTGGGCAATGTGTGCTTCTCTCAGAGTGAACAGAGCC 120
 Qy 975 GCTGTACCCAGTGCAGGTCACTGTGGAGACGCTCGGCTCATGTCTTTGACAAGAGCGA 1034
 Db 121 GCTGTACCCAGTGCAGGTCACTGTGGAGACGCTCGGCTCATGTCTTTGACAAGAGCGA 180
 Qy 1035 AGGAGGTGGGGGCTGTGTGCTCTCGGCTCAACAGCCAGGGTACCGGACTCAGCTG 1094
 Db 181 AGGAGGTGGGGGCTGTGTGCTCTCGGCTCAACAGCCAGGGTACCGGACTCAGCTG 240
 Qy 1095 CGAGAGATGGGCTTCTCAGGGGCACTGACCACTCCGAGCTGGACGTGGCAACGGGGG 1154

Db 643 CACCTGTGTGGGGATCCCTGCTCTCCGGGGAGTGGGTTGTGACAGCCGCCACCTGCTTC 702

Qy 1441 CCGGAGCGGAACCGGGTCTCTGCCCCGATGGGAGTGTGGCGGTGCGGTGCCCAAGCC 1500

Db 703 CCGGAGCGGAACCGGGTCTCTGCCCCGATGGGAGTGTGGCGGTGCGGTGCCCAAGCC 762

Qy 1501 TCTCCCAACGCTGACAGCTGGGGGTGAGGCTGTGCTACCAAGGGGGCTATCTTCC 1560

Db 763 TCTCCCAACGCTGACAGCTGGGGGTGAGGCTGTGCTACCAAGGGGGCTATCTTCC 822

Qy 1561 TTTCGGGACCCCAACAGGAGGAGAAACGACGATATGCTGCTGCTCCACTCTCCAGT 1620

Db 823 TTTCGGGACCCCAACAGGAGGAGAAACGACGATATGCTGCTGCTCCACTCTCCAGT 882

Qy 1621 CCCCCTGCCCCCTCAGAAATATATCCAGCTGTGTGCTCCCAAGCTGCGGGCAGGCTCTG 1680

Db 883 CCCCCTGCCCCCTCAGAAATATATCCAGCTGTGTGCTCCCAAGCTGCGGGCAGGCTCTG 942

Qy 1681 GTGATGGCAAGATCTGTACCGGTGACGGGCTGGGGCAACGACGATCTATGCGCAACAG 1740

Db 943 GTGATGGCAAGATCTGTACCGGTGACGGGCTGGGGCAACGACGATCTATGCGCAACAG 1002

Qy 1741 GCCGGGGTACTCCAGAGGCTCGAGTCCCAATATCAGCAATGATGTCTGCAATGGCGCT 1800

Db 1003 GCCGGGGTACTCCAGAGGCTCGAGTCCCAATATCAGCAATGATGTCTGCAATGGCGCT 1062

Qy 1801 GACTTTTATGGAACCAAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAAGGTGGC 1860

Db 1063 GACTTTTATGGAACCAAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAAGGTGGC 1122

Qy 1861 ATTGATGCTTCCAGAGGCGGAGCGGAGTGGTCCCTTTGTGTGAGAGACAGATCTCTCG 1920

Db 1123 ATTGATGCTTCCAGAGGCGGAGCGGAGTGGTCCCTTTGTGTGAGAGACAGATCTCTCG 1182

Qy 1921 ACGCCAGCTTGGCGGCTGTGTGAGATTTGAGATTGGGGCACTGCTGTGCTCCGCGCCAG 1980

Db 1183 ACGCCAGCTTGGCGGCTGTGTGAGATTGGGGCACTGCTGTGCTCCGCGCCAG 1242

Qy 1981 AAGCCAGGCGCTTCAACCAAGTCAAGTACTTCCGGAGTGGATCTTCCAGGCCATTA-- 2038

Db 1243 AAGCCAGGCGCTTCAACCAAGTCAAGTACTTCCGGAGTGGATCTTCCAGGCCATTAAG 1302

Qy 2039 ----- 2038

Db 1303 GTGAAAGTTGGGTCCAGATGGGACCAAGGTGGGACGTTGGGTCTTAATGGGGGAAG 1362

Qy 2039 ----- 2038

Db 1363 GGAGGAGAGATTTGTTTAGAAACTACGCTCAGGCTAGAGAGGCGCCCTTGAG 1422

Qy 2039 ----- 2038

Db 1423 AACAGATGACTTTGAAGGTTCTCTGGGAAAGGAAGCAAGTGTGGACGTGAAGCT 1482

Qy 2039 ----- AGACTCACTCCGAAGCAGCGGCGATG 2064

Db 1483 CTGACAGCTGGGAGCCCCCAGCTGTCTTTTCCCAAGACTCACTCCGAAGCCAGCGGCGATG 1542

Qy 2065 GTGACCCAGCTCTGAACCGGTGCTTCTGTGCTGCGAGGCTTCAGGGCCCGAGGTGATCC 2124

Db 1543 GTGACCCAGCTCTGAACCGGTGCTTCTGTGCTGCGAGGCTTCAGGGCCCGAGGTGATCC 1602

Qy 2125 GGTGTGGGATCCACGCTGGGCGGAGATGGGAGCTTTTCTTCTTGGGCGCGGTCCACA 2184

Db 1603 GGTGTGGGATCCACGCTGGGCGGAGATGGGAGCTTTTCTTCTTGGGCGCGGTCCACA 1662

Qy 2185 GGTGCAAGGACACCTCCCTCAGGGTCTCTCTTCCAGTGGGCGGCGCACTCAGGCC 2244

Db 1663 GGTGCAAGGACACCTCCCTCAGGGTCTCTCTTCCAGTGGGCGGCGCACTCAGGCC 1722

Qy 2245 CGAGACCAACCACTCTCTCTGACGCCCATATGAATATGTTCTGTGCTGTGGAC 2304

Db 1723 CGAGACCAACCACTCTCTCTGACGCCCATATGAATATGTTCTGTGCTGTGGAC 1782

Qy 2305 TCTGTAGTGGCCCTGATGATGGGATGCTCTTAAATATATAA 2350

Db 1783 TCTGTCTAGGTGCTCTGATGAGGAGTCTCTTAAATATATAA 1828

RESULT 4
AK002694
LOCUS
DEFINITION
MUS musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610030A17 product:hepsin, full insert sequence.
ACCESSION
AK002694
VERSION
AK002694.1 GI:12832864
KEYWORDS
HTC; CAP trapper.
SOURCE
MUS musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
Carninci, P., and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
PUBMED
99279253
REFERENCE
AUTHORS
1
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
PUBMED
11042159
REFERENCE
AUTHORS
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
PUBMED
11042159
REFERENCE
AUTHORS
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Maruyama, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL
Genome Res. 10 (11), 1757-1771 (2000)
PUBMED
20530913
REFERENCE
AUTHORS
4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE
Functional annotation of a full-length mouse cDNA collection
JOURNAL
Nature 409, 685-690 (2001)
PUBMED
11076861
REFERENCE
AUTHORS
5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
PUBMED
12111111
REFERENCE
AUTHORS

TITLE
JOURNAL
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

Db 1591 GGCCTCACACTGAGGCTCCATCATGGAATGTTTCTGCTCAGATCCAGTCCAGGGGTCCAA 1650
QY 2192 GGAACACCTCCCTCCAGGGTCTCTCTTCCACAGTGGGGGGCCCACTAGCCCCGAGACC 2251
Db 1651 GG--ATGCTGATCCAGGAGCTTCTCTTCCACAGGGCCGCCCACTCAATCCAGGGCC 1708
QY 2252 ACCCACTCAGCCCTCCAGGCCCCCATGTAATATATTTGTTGCTGCTGGAGACTCTGTC 2311
Db 1709 ATTGGCTCACCCTCC--ACCCATGTAATATATTTACTCTGCTCTGGGGGGGCGCTCT 1765
QY 2312 TAGGTGCCCCCTGATGATGGATGCTCTTTAAATAATAAAGATGTTTT 2359
Db 1766 AGGAGAGCCCCCT--TGTGCAATGCTCTTAAATAATAAAGGTGTTTT 1811

RESULT 5
AY408326 1238 bp DNA linear GSS 15-DEC-2003
LOCUS Mus musculus HPN gene, VIRtual TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY408326
VERSION AY408326.1 GI:39764297
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Eutelestomi;
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
1 (bases 1 to 1238)
Todd,M.A., Tenenbaum,D.M., Ciavella,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1238)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tenenbaum,D.M., Ciavella,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1. .1238
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
gene
<1..>1238
/gene="HPN"
/locus_tag="HGM3186"

ORIGIN
Query Match 39.5%; Score 932.4; DB 9; Length 1238;
Bee local Similarity 84.6%; Pred. No. 3.2e-206;
Matches 1047; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

QY 842 GTGGCCGAGCTGTGCATGCTGCTCCAGACCAAGTGGAGCTCTCACTCGGGAGACC 901
Db 1 GTGGCCGAGCTGTGCATGCTGCTCCAGACCAAGTGGAGCTCTCACTCGGGAGACC 60
QY 902 TGCTACTTGTGACAGCCATCGGGCGGAGCTCTGGGCCATTGTGGCTGTTCTCTCAGA 961
Db 61 TGCTACTTGTGACAGCCATCGGGCGGAGCTCTGGGCCATTGTGGCTGTTCTCTCAGA 120
QY 962 GTGACCGAGAGCCGCTGTACCAAGTGCAGTCAAGTCTGCGGAGCGCTGGCTCATGCTCT 1021
Db 121 GTGACCGAGAGCCGCTGTACCAAGTGCAGTCAAGTCTGCGGAGCGCTGGCTCATGCTCT 180
QY 1022 TTGACAAGAGCGAAGGAGCTGTGGCGGCTGTGCTCTCGCGCTCCAAAGCCAGGGTAG 1081

Db 181 TTGACAAGAGCGAAGGAGCTGTGAGGCTGCTGCTCTCAAGCTCCAAAGTCCAGGGTAG 240
QY 1082 CCGGACTGAGCTGCGAGAGATGGGCTTCTCCAGGGCACTGACCACTCCGAGCTGAGAG 1141
Db 241 CAGGCTCTGCTGTGAGAGATGGGCTTCTCCAGGGCTCTGGCACTCTCGAGCTGAGATG 300
QY 1142 TTGCAAGGGGGGGGCGCAATGAGCAAGTGGGCTTCTCTGATGAGAGAGGGAGGCTGC 1201
Db 301 TGCCCACTGGGGGGCGCAAGGCAATGGGCTTCTTGTGATGAGAGAGGGCGGACTTC 360
QY 1202 CCCAACCAGAGAGCTGCTGAGAGTCACTCCGATGATGATGATGATGATGATGATGATGAT 1261
Db 361 CTCTGCTCAGAGGTTCTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 1262 TGAGCCCATCTGCGAAGACTGTGGCCCGAGAGAGCTGCCGATGATGATGATGATGATGAT 1321
Db 421 TGACTGCTCAGCTGCGAAGACTGTGGCCCGAGAGAGCTGCCGATGATGATGATGATGATGAT 480
QY 1322 GCGGAGACACAGCTTGGGGCGGAGGCGGATGAGGAGCTGAGGAGCTTGGATGATGATGATGAT 1381
Db 481 GCGAGAGAGAGCTGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
QY 1382 ACTCTGTGGGGAGATCCCTGCTCTCCGGGAGCTGGAGCTGAGACAGCCGCACTGCTTC 1441
Db 541 ACTCTGTGGGGAGATCCCTGCTCTCCGGGAGCTGGAGCTGAGACAGCTGAGACATGCTTTC 600
QY 1442 CGAGCGGAGAACCGGCTCTCTCCGATGAGGAGGATGTTGGCGGCTGCGGCTGAGGCT 1501
Db 601 CAGAGCGGAGAACCGGCTCTCTCTCCGATGAGGAGGATGTTGGCGGCTGAGGCTGAGGCT 660
QY 1502 CTCCCAAGGCTGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1561
Db 661 CACCCATGCTGTGAGCACTGGGGGCTGAGGCTGATCTATGAGGGGCTTACTTCCCT 720
QY 1562 TTGCGGAGCCCAACAGGAG 1621
Db 721 TTGAGAGCCCTGATGAG 780
QY 1622 CCTGCGCTTCACAGATATCATTCAGGCTGTGAGCTCCAGAGCTGCGGCGGAGGCTTGG 1681
Db 781 CCTGCGCTTCACAGATATCATTCAGGCTGTGAGCTCCAGAGCTGCGGCGGAGGCTTGG 840
QY 1682 TTGATGCGAAGATGTGATGAGCTGAGCGGCTGGGGGAGAGAGAGAGAGAGAGAGAGAG 1741
Db 841 TTGATGCGAAGATGTGATGAGCTGAGCGGCTGGGGGAGAGAGAGAGAGAGAGAGAGAG 900
QY 1742 CCGGGTACTTCAGAGAGCTGAGTCCCATATATCAGCAATGATGTCTGCAATGGCGCTG 1801
Db 901 CTATGAGTCTCCAAAGAGGCGGAGTCCCATATATCAGCAATGATGTCTGCAATGGCGCTG 960
QY 1802 ACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCTGCTGCTGCTGCTGCT 1861
Db 961 ACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCTGCTGCTGCTGCTGCT 1020
QY 1862 TTGATGCTGCGAAGGAG 1921
Db 1021 TTGATGCTGCGAAGGAG 1080
QY 1922 CGCCACGTTGCGGCTGTGTGAGCAATGTGATGATGATGATGATGATGATGATGATGATGAT 1981
Db 1081 CATCAAGTGGCGGCTATGTGAGCAATGTGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1982 AGCCAGGCTGTACCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2041
Db 1141 AGCCAGGAGTGTACCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1200
QY 2042 CTCACTCGAAGCGAGCGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2079
Db 1201 CTCACTCGAAGCGAGCGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1238

RESULT 6

CR592189 1212 bp mRNA linear HNC 21-JUL-2004
 LOCUS full-length cDNA clone CS0DM012Y015 of Fetal Liver of Homo sapiens (human)
 ACCESSION CR592189
 VERSION CR592189.1 GI:50472996
 KEYWORDS HTC; cNSLT cDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1212)
 AUTHORS Li, W.B., Gruber, C., Jeese, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 CONTACT : Peng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ invitrogen Corporation 1600 Paraday Avenue
 REMARK 2 (bases 1 to 1212)
 REFERENCE Genoscope.
 AUTHORS Direct Submission
 TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
 JOURNAL - Web : www.genoscope.cns.fr)
 COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
 FEATURES
 source Location/Qualifiers
 1..1212
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DM012Y015"
 /tissue_type="Fetal liver"
 /plasmid="pCMVSPORT_6"

ORIGIN
 Query Match 38.3%; Score 905.4; DB 3; Length 1212;
 Best Local Similarity 99.3%; Pred. No. 6.2e-200;
 Matches 909; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

1336 GCTTCCCGAGCGAACCAGGCTCTGTCTCCGATGCGAGTGTTCGCGCTGCGCC 1495
 298 GCTTCCCGAGCGAACCAGGCTCTGTCTCCGATGCGAGTGTTCGCGCTGCGCC 357
 1496 AGGCTCTCCCGAGCGAACCAGGCTCTGTCTCCGATGCGAGTGTTCGCGCTGCGCC 1555
 358 AGGCTCTCCCGAGCGAACCAGGCTCTGTCTCCGATGCGAGTGTTCGCGCTGCGCC 417
 1556 TTCCCTTCGGGAGCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1615
 418 TTCCCTTCGGGAGCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 477
 1616 CCAAGTCCCTGCGCTCAAGAGATTCATCCAGCTGTGTGCTTCGAGCTGCGCGCAGG 1675
 478 CCAAGTCCCTGCGCTCAAGAGATTCATCCAGCTGTGTGCTTCGAGCTGCGCGCAGG 537
 1676 CCCTGGTGGATGGCAAGATCTGTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1735
 538 CCCTGGTGGATGGCAAGATCTGTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 597
 1736 AACAGGCGGGGATCTCCAGAGAGGCTGAGTCCCATATATCAGCATATATGTCTGCAATG 1795
 598 AACAGGCGGGGATCTCCAGAGAGGCTGAGTCCCATATATCAGCATATATGTCTGCAATG 657
 1796 GCGGTGATTTCTATGGAACCAAGATCAAGCCCAAGATTTCTGTGCTGGCTACCCCGAGG 1855
 658 GCGGTGATTTCTATGGAACCAAGATCAAGCCCAAGATTTCTGTGCTGGCTACCCCGAGG 717
 1856 GTGCGATGATGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1915

Db 718 GTGCGATGATGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATCT 777
 Qy 1916 CTGCGAGCGCACGCTGGCGGCTGTGTGCGATTTGAGTTGGGAGCATGCGCTGTGCGCTGG 1975
 Db 778 CTGCGAGCGCACGCTGGCGGCTGTGTGCGATTTGAGTTGGGAGCATGCGCTGTGCGCTGG 837
 Qy 1976 CCGAAGGCGGAGGAGGCTGTACACCAAGGACAGTCTCCGGAGGTGATCTTCCAGAGCA 2035
 Db 838 CCGAAGGCGGAGGAGGCTGTACACCAAGGACAGTCTCCGGAGGTGATCTTCCAGAGCA 897
 Qy 2036 TAAAGACTCACTCCGAAGCCAGCGGATGTGACCCAGCTTGAACCGGTGCTTCTGCT 2095
 Db 898 TAAAGACTCACTCCGAAGCCAGCGGATGTGACCCAGCTTGAACCGGTGCTTCTGCT 957
 Qy 2096 GCGAGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2155
 Db 958 GCGAGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1017
 Qy 2156 GACGTTTTTCTTTGTTGGGCGGCTGTCACAGGTCCAGAGACACCTCCCTCCAGAGGCTCTC 2215
 Db 1018 GACGTTTTTCTTTGTTGGGCGGCTGTCACAGGTCCAGAGACACCTCCCTCCAGAGGCTCTC 1077
 Qy 2216 TCTTCCAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2275
 Db 1078 TCTTCCAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1137
 Qy 2276 CATGTAATATATTGTTGCTGTCTGTGAGGATCTCTGTGAGGAGGAGGAGGAGGAGGAGG 2335
 Db 1138 CATGTAATATATTGTTGCTGTCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1197
 Qy 2336 TCTTAAATATATAA 2350
 Db 1198 TCTTAAATATATAA 1212

RESULT 7
 AL558100 846 bp mRNA linear EST 02-APR-2004
 LOCUS AL558100 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
 DEFINITION Homo sapiens cDNA clone CS0DJ002YF16 5-PRIME, mRNA sequence.
 ACCESSION AL558100
 VERSION AL558100.3 GI:46183500
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 846)
 AUTHORS Li, W.B., Gruber, C., Jeese, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On Feb 15, 2001 this sequence version replaced gi:31279899.
 CONTACT : Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Creteil, CP 5706 - 91057 Evry cedex - FRANCE
 Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9952.f
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?b=CS0DJ002D08P1&c=9952.f.
 FEATURES
 source Location/Qualifiers
 1..846
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DJ002YF16"
 /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /cell_line="JURKAT"
 /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT

10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 35.2%; Score 811; DB 1; Length 846;
Best Local Similarity 99.4%; Pred. No. 1.3e-182;
Matches 842; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 97 GCTGCTGGGGGCCACATGCTCTCTGCGCCAGGCTTGAAGTACCGGACCCCGCACTA 156
DB 1 GCTGCTGGGGGCCACATGCTCTCTGCGCCAGGCTTGAAGTACCGGACCCCGCACTA 60
QY 157 CCTGAGGCTCCGGCCCGCCAGCTGAGACCCGAGGTAAAGACAAAGGGCCCGCAGACTCA 216
DB 61 CCTGAGGCTCCGGCCCGCCAGCTGAGACCCGAGGTAAAGACAAAGGGCCCGCAGACTCA 120
QY 217 CAGTTCCAGCCCTGAGAGACAGAGGGTTCCTCATCCCGCCAGCCCTAATAGCCCACTC 276
DB 121 CAGTTCCAGCCCTGAGAGACAGAGGGTTCCTCATCCCGCCAGCCCTAATAGCCCACTC 180
QY 277 CTAATAGAGGGGTTCTGAGGACCTGAAGAGGGGACTATAGAGTCTCCCAAGACTT 336
DB 181 CTAATAGAGGGGTTCTGAGGACCTGAAGAGGGGACTATAGAGTCTCCCAAGACTT 240
QY 337 AGGATCTCTGCT 396
DB 241 AGGATCTCTGCT 300
QY 397 AGACCCAGAGATTCAGGCTCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 456
DB 301 AGACCCAGAGATTCAGGCTCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
QY 457 AAATTCCT 516
DB 361 AAATTCCT 420
QY 517 GGGATCTCTGCT 576
DB 421 GGGATCTCTGCT 480
QY 577 CAGGCTCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 636
DB 481 CAGGCTCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 637 GTCCGTGATCAGGCGTGTCTGTGCGCAAGGCCAGTCTCTAAGCTGTGCTGTGATGAGC 696
DB 541 GTCCGTGATCAGGCGTGTCTGTGCGCAAGGCCAGTCTCTAAGCTGTGCTGTGATGAGC 600
QY 697 CCTGAGGCTCCGGCCCGCCAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 756
DB 601 CCTGAGGCTCCGGCCCGCCAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 659
QY 757 TCCATCTCTCTCAGAGTCTCCACCTGTGCGCCAGAGTCTAGCCAGGAAATCTTAACAGA 816
DB 660 TCCATCTCTCAGAGTCTCCACCTGTGCGCCAGAGTCTAGCCAGGAAATCTTAACAGA 719
QY 817 GGCAGTACATGCGCAGAGAGAGGGTGTGCGGACTGTGCTGCTCTCAAGACCAAG 876
DB 720 GGCAGTACATGCGCAGAGAGAGGGTGTGCGGACTGTGCTGCTCTCAAGACCAAG 779
QY 877 GTGACAGCTCAGTGGCGGGGACCTGTGCTAATTCTGACAGGCAATCGGGGGGGGACCTCGG 936
DB 780 GTGACAGCTCAGTGGCGGGGACCTGTGCTAATTCTGACAGGCAATCGGGGGGGGACCTCGG 839
QY 937 GCCATTG 943
DB 840 GGCAATTG 846

RESULT 8

BG754441
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BG754441 916 bp mRNA linear EST 15-MAY-2001
60271007F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4846387 5',
mRNA sequence.
BG754441
BG754441.1 GI:14065106
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NIH-MGC <http://mhc.nci.nih.gov/>.
1 (bases 1 to 916)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
<http://image.lmnl.gov>
Plate: LCM1687 row: b column: 20
High quality sequence stop: 803.
Location/Qualifiers
1. 916
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4846387"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_48"
/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

FEATURES

source

ORIGIN

Query Match 34.4%; Score 814; DB 4; Length 916;
Best Local Similarity 95.8%; Pred. No. 1.2e-178;
Matches 879; Conservative 0; Mismatches 35; Indels 4; Gaps 4;

QY 1026 CAAGACGGAAGGACGTGGCGGCTGTCTCTCTCGCGCTCCAAAGCCAGGGTAGCCGG 1085
DB 2 CAAGACGGAAGGACGTGGCGGCTGTCTCTCTCTCGCGCTCCAAAGCCAGGGTAGCCGG 60
QY 1086 ACTCAGCTGCGAGAGATGGGCTTCTCTCAGGAGCACTGACCCTCGAGCTGACGTGCG 1145
DB 61 ACTCAGCTGCGAGAGATGGGCTTCTCTCAGGAGCACTGACCCTCGAGCTGACGTGCG 120
QY 1146 AAGCGCGGGGCCAATGACACGTGCGGCTTCTCTGTGTGAGAGAGAGGGGAGGCGCCCA 1205
DB 121 AAGCGCGGGGCCAATGACACGTGCGGCTTCTCTGTGTGAGAGAGAGGGGAGGCGCCCA 180
QY 1206 CACCCAGAGGCTGTGAGGTCATCTCTCGTGTGATTTGCCCGCAGAGGCGCTTCTTGAGC 1265
DB 181 CACCCAGAGGCTGTGAGGTCATCTCTCGTGTGATTTGCCCGCAGAGGCGCTTCTTGAGC 240
QY 1266 CGGCATCTGCGAAGCTGTGGCGCAGAGAGTGCCTCGTGAACCGCATGCTGAGAGCGC 1325
DB 241 CGGCATCTGCGAAGCTGTGGCGCAGAGAGTGCCTCGTGAACCGCATGCTGAGAGCGC 300
QY 1326 GGAACCAAGCTTGGGCGGCTGGCGGAGAGTCAAGTCAAGCTTCTGATGATGAGACACACT 1385
DB 301 GGAACCAAGCTTGGGCGGCTGGCGGAGAGTCAAGTCAAGCTTCTGATGATGAGACACACT 360

1386 CTGTGGGGGATCCCTGCTCTCCGGGAGTGGTGTCTGAACAGCCGCCCTGCTTCCCGGA 1445
DB 361 CTGTGGGGGATCCCTGCTCTCCGGGAGTGGTGTCTGAACAGCCGCCCTGCTTCCCGGA 420
1446 GCGGAACCGGGTCTGCTCCGATGGAGTGTTCGCGGTGCGGCGGAGGCTCTCC 1505
DB 421 GCGGAACCGGGTCTGCTCCGATGGAGTGTTCGCGGTGCGGCGGAGGCTCTCC 480
1506 CCACGCTCTGAGCTGGGGGCTGACAGCTGTGTCTACCAAGGGGCTATCTTCCCTTTCG 1565
DB 481 CCACGCTCTGAGCTGGGGGCTGACAGCTGTGTCTACCAAGGGGCTATCTTCCCTTTCG 540
1566 GGACCCCAACAGCGAGGAGAACAGACGATATTGCTCTGCTTCCAGTCTCTCAAGTCCCT 1625
DB 541 GGACCCCAACAGCGAGGAGAACAGACGATATTGCTCTGCTTCCAGTCTCTCAAGTCCCT 600
1626 GCGCCCTCAAGATATCATCTCAAGCTGTGTGCTCCAGCTGCGGCGAGGCTGTGTGA 1685
DB 601 GCGCCCTCAAGATATCATCTCAAGCTGTGTGCTCCAGCTGCGGCGAGGCTGTGTGA 660
1686 TGGCAAGATCTGTACCGTGTGAGCGGCTGGGCAACAGCAATATATGCAACAGCCGG 1745
DB 661 TGGCAAGATCTGTACCGTGTGAGCGGCTGGGCAACAGCAATATATGCAACAGCCGG 720
1746 GGTACTCCAGAGGCTCGAGTCCCATATATCAAGCAATATATGCTGTGCAATGCGCTGACTT 1805
DB 721 GGTACTCCAGAGGCTCGAGTCCCATATATCAAGCAATATATGCTGTGCAATGCGCTGACTT 780
1806 CTATGGAACAGATCAAGCCCAAGATGTCTGTGCTGCTGCTACCCCGAGGCTGATGA 1865
DB 781 CTATGGAACAGATCAAGCCCAAGATGTCTGTGCTGCTGCTACCCCGAGGCTGATGA 838
1866 TGCCTGCGAGGCGGCAAGCGGTGTCTCTTGTGTGTGAGAGCAAGCATCTCTCGAGCC 1925
DB 839 TGCCTGCGAGGCGGCGGAGGTGTCTCTTGTGTGTGAGAGCAAGCATCTCTCGAGCC 897
1926 ACCTTGGCGGCTGTGTGTG 1943
DB 898 TTGCGGGTGTGTGTGTG 915

RESULT 9

LOCUS

BG754531

DEFINITION BG754531 940 bp mRNA linear EST 15-MAY-2001
602710173F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4846386 5',
mRNA sequence.

ACCESSION

BG754531

VERSION

BG754531.1

GI:14065197

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 940)

TITLE

NIH-MGC http://mgi.nci.nih.gov/

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgabbs-rc@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCM1687 row: b column: 19
High quality sequence stop: 833.
Location/Qualifiers
1..940
/organism="Homo sapiens"
/mol_type="mRNA"

/db_xref="taxon:9606"
/clone="IMAGE:4846386"
/issue_type="Primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_48"
/note="Organ: B-cells; Vector: pORF7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Query Match 33.7%; Score 796.8; DB 4; Length 940;

Best Local Similarity 96.7%; Pred. No. 1.2e-174;

Matches 888; Conservative 0; Mismatches 22; Indels 8; Gaps 7;

1026 CAAGACGAGAGGACGTCGCGGCTGTGTCTCTCCGCTTCCAAACCGAGGTTACCCG 1085
DB 2 CAAGACGAGAGGACGTCGCGGCTGTGTCTCTCCGCTTCCAAACCGAGGTTACCCG 60
1086 ACTGAGCTGCGAGGAGATGGGCTTCTCAGGGCACTGACCCGAGCTGGAAGTGG 1145
DB 61 ACTGAGCTGCGAGGAGATGGGCTTCTCAGGGCACTGACCCGAGCTGGAAGTGG 120
1146 AACGCGGGCGCCAAATGACGTCGCGGCTTCTGTGTGAGACAGAGGAGGCTGCCCA 1205
DB 121 AACGCGGGCGCCAAATGACGTCGCGGCTTCTGTGTGAGACAGAGGAGGCTGCCCA 180
1206 CACCAAGAGCTGTGAGAGTATCTCTGCTGTGTGATTTGCCAGAGGCGCTTCTTGGC 1265
DB 181 CACCAAGAGCTGTGAGAGTATCTCTGCTGTGTGATTTGCCAGAGGCGCTTCTTGGC 240
1266 CGCCATCTGCCAAGCTGTGCGGCGGAGAGCTCCGCTGGAACCGCATCTGGAGAGCG 1325
DB 241 CGCCATCTGCCAAGCTGTGCGGCGGAGAGCTCCGCTGGAACCGCATCTGGAGAGCG 300
1326 GGAACACGCTTGGGCGGCTGAGCGGCTGAGCAAGTCAAGCTTGTGATGAGACACACT 1385
DB 301 GGAACACGCTTGGGCGGCTGAGCGGCTGAGCAAGTCAAGCTTGTGATGAGACACACT 360
1386 CTGTGGGGGATCCCTGCTCTCCGGGAGCTGAGGCTGAGACAGCCGCCCTTCCCGGA 1445
DB 361 CTGTGGGGGATCCCTGCTCTCCGGGAGCTGAGGCTGAGACAGCCGCCCTTCCCGGA 420
1446 GCGGAACCGGGTCTGTGTCCGATGGCGAGTCTTGGCGGCTGAGCGGCTGAGCGGCT 1505
DB 421 GCGGAACCGGGTCTGTGTCCGATGGCGAGTCTTGGCGGCTGAGCGGCTGAGCGGCT 480
1506 CCACGCTCTGAGCTGGGGGCTGAGGCTGTGTGCTTACCAAGGGGCTATCTTCCCTTTCG 1565
DB 481 CCACGCTCTGAGCTGGGGGCTGAGGCTGTGTGCTTACCAAGGGGCTATCTTCCCTTTCG 540
1566 GGACCCCAACAGCGAGGAGAACAGCAAGATATTGCTGTGCTCAAGCTTCCAGTCCCT 1625
DB 541 GGACCCCAACAGCGAGGAGAACAGCAAGATATTGCTGTGCTCAAGCTTCCAGTCCCT 600
1626 GCGCCCTCAAGATATCATCTCAAGCTGTGTGCTCCAGCTGCGGCGAGGCTGTGTGA 1685
DB 601 GCGCCCTCAAGATATCATCTCAAGCTGTGTGCTCCAGCTGCGGCGAGGCTGTGTGA 660
1686 TGGCAAGATCTGTACCGTGTGAGCGGCTGGGCAACAGCAATATATGCAACAGCCGG 1745
DB 661 TGGCAAGATCTGTACCGTGTGAGCGGCTGGGCAACAGCAATATATGCAACAGCCGG 719
1746 GGTACTCCAGAGGCTGAGTCCCATATATCAAGCAATATATGCTGTGCAATGCGCTGACTT 1805
DB 720 GGTACTCCAGAGGCTGAGTCCCATATATCAAGCAATATATGCTGTGCAATGCGCTGACTT 778
1806 CTATGGAACAGATCAAGCCCAAGATGTCTGTGCTGAGTACCCCAAGGCTGACTTG 1864

Db 779 CTATGAAACC--GATCAAGCCAGATGTTCTGTGCTCCCTGATACCCCGAGGGGTGCTTTG 836
 Qy 1865 ATGCTTCGACAGGGGACAGCCGCTGCTCCCTTTGTGTGTAGACACAGCATCTCTTGAGACG 1924
 Db 837 ATTCTGTGACAGGGGACCGCGGTGCT--CTTTGTGTGTAGAGCCGAC--TCTCTGAGACG 894
 Qy 1925 CAGCTTGGCGGCTGTGTG 1942
 Db 895 CCGTGGCGGCTGTGTG 912
 RESULT 10
 BX406458 965 bp mRNA linear EST 01-MAY-2004
 LOCUS BX406458
 DEFINITION BX406458 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
 ACCESSION CS0DM011Y04 5-PRIME, mRNA sequence.
 VERSION BX406458
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 AUTHORS Full-length cDNA libraries and normalization
 TITLE Unpublished (2001)
 JOURNAL Contact: Genoscope
 COMMENT On May 13, 2003 this sequence version replaced gi:30644262.
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
 was not normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 This sequence belongs to sequence cluster 9952.f
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?S=CS0AM011DB02Q1&c=9952.f.
 Location/Qualifiers
 1..965
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DM011Y04"
 /library="FETAL LIVER"
 /dev_stage="fetal"
 /clone_lib="Homo sapiens FETAL LIVER"
 /notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
 was primed with a NotI-oligo (dT) primer. Five prime end
 enriched, double-strand cDNA was digested with Not I and
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6
 vector. Library was not normalized."
 ORIGIN
 Query Match 33.6%; Score 794; DB 5; Length 965;
 Best Local Similarity 93.1%; Pred. No. 5.5e-174;
 Matches 833; Conservative 16; Mismatches 44; Indels 2; Gaps 2;
 Qy 729 CTGGGCTCCCGCAGGCGCTGCTCCCGCTCATCTCCCAAGAGTCCCAAGCTGAGCCAG 788
 Db 71 CGGCACTACCTCGAGGCTCCGCCCCACCTGCTGAGACCCAGAGGTCACCTGAGCCAG 130
 Qy 789 GAGGTCAAGCCAGGGAATCATTAACAAGAGGAGTGCATGCGGAGAGAGAGGAGGTCGCG 848
 Db 131 GAGGTCAAGCCAGGGAATCATTAACAAGAGGAGTGCATGCGGAGAGAGAGGAGGTCGCG 190
 Qy 849 GACTGTGCATGCTGCTTCAAGACCAAGAGTGGCAGCTTCACTGGGGGAGACCTGTACT 908
 Db 191 GACTGTGCATGCTGCTTCAAGACCAAGAGTGGCAGCTTCACTGGGGGAGACCTGTACT 250

Qy 909 TCTGACAGCATCGGGGCGGATCTCTGAGCCATTGTGGATTTCTCTCAGAGATGACA 968
 Db 251 TCTGACAGCATCGGGGCGGATCTCTGAGCCATTGTGGATTTCTCTCAGAGATGACA 310
 Qy 969 GGAAGCCGCTGTACCAATGACAGTCAAGCTTGCAGGAGCTGTGAGCTTGTGACAA 1028
 Db 311 GGAAGCCGCTGTACCAATGACAGTCAAGCTTGCAGGAGCTGTGAGCTTGTGACAA 370
 Qy 1029 GACGGAAGGACCGTGGGCGCTGTGCTCTCTGCGCTCCAAAGCCAGGGAGCGAGCT 1088
 Db 371 GACGGAAGGACCGTGGGCGCTGTGCTCTCTGCGCTCCAAAGCCAGGGAGCGAGCT 430
 Qy 1089 CAGCTGCGAGAGATGAGGCTTCTCAGAGGACCTGACCCAGCTCCAGCTGACGTCGAA 1148
 Db 431 CAGCTGCGAGAGATGAGGCTTCTCAGAGGACCTGACCCAGCTCCAGCTGACGTCGAA 490
 Qy 1149 GCGGCGCGCAATGACAGTGGGCTTCTTGTGTGTGACGAGGAGAGCTGCCACAC 1208
 Db 491 GCGGCGCGCAATGACAGTGGGCTTCTTGTGTGTGACGAGGAGAGCTGCCACAC 550
 Qy 1209 CCAGAGGCTGCTGAGGCTCATCTCCGCTGTGATGATGATGATGATGATGATGATGAT 1268
 Db 551 CCAGAGGCTGCTGAGGCTCATCTCCGCTGTGATGATGATGATGATGATGATGATGAT 610
 Qy 1269 CATCTGCGAAGACTGTGGCGCAGAGAGCTGCCGCTGAGACCGCATCTGTGGAGGCGGGA 1328
 Db 611 CATCTGCGAAGACTGTGGCGCAGAGAGCTGCCGCTGAGACCGCATCTGTGGAGGCGGGA 669
 Qy 1329 CACCAAGTTGGGCGCGTGGCGGTCGAGAGTGAAGCTTGTGATGATGATGATGATGAT 1388
 Db 670 CACCAAGTTGGGCGCGTGGCGGTCGAGAGTGAAGCTTGTGATGATGATGATGATGAT 728
 Qy 1389 TGGGGGATCCCTGCTGCTCCGGGAGCTGGGTGCTGACAGCGCCGACCTGTTCCGGAGCG 1448
 Db 729 TGGGGGATCCCTGCTGCTCCGGGAGCTGGGTGCTGACAGCGCCGACCTGTTCCGGAGCG 788
 Qy 1449 GAACCGGAGTCTGCTCCGATGAGTGGCGAGTGTGCGGAGCTGGCGCAGAGCTTCCCA 1508
 Db 789 GAACCGGAGTCTGCTCCGATGAGTGGCGAGTGTGCGGAGCTGGCGCAGAGCTTCCCA 848
 Qy 1509 CGGTCTGACGTGGGGGTGACAGCTGTGTGCTTACACAGGGGCTATCTTCCCTTGGGA 1568
 Db 849 CGGTCTGACGTGGGGGTGACAGCTGTGTGCTTACACAGGGGCTATCTTCCCTTGGGA 908
 Qy 1569 CCCCAGAGGAGAGAGAGAGAGAGATTTGCTCCCTGCTCCAGCTCCAGTCCC 1623
 Db 909 CCCCAGAGGAGAGAGAGAGAGATTTGCTCCCTGCTCCAGCTCCAGTCCC 963
 RESULT 11
 CK231686
 LOCUS CK231686 1012 bp mRNA linear EST 09-DEC-2003
 DEFINITION ILUMIGEN MCQ_2776 Katze_MtLg Macaca mulatta cDNA 5' similar to
 human HPN (hs.432750), mRNA sequence.
 ACCESSION CK231686
 VERSION CK231686.1 GI:39638044
 KEYWORDS EST.
 SOURCE Macaca mulatta (rhesus monkey)
 ORGANISM Macaca mulatta
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.
 REFERENCE Katze, M.G., Thomas, M., Korth, M., Iadonato, S.P. and Magness, C.L.
 AUTHORS Large-scale Rhesus Macaque cDNA Sequencing
 TITLE Unpublished (2003)
 JOURNAL Contact: C. Magness
 COMMENT Illumigen Biosciences Inc.
 2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
 Tel: 2063780400
 Fax: 2063780408
 Email: cmagness@illumigen.com
 Sequenced on 2003.11.05. 744 Q20 bases. Assemblies in contig w/ 2

798 CAGGGAATCATTAACAAGAGGCACTGACATGGCCAGAGAAGGCTGCCCGGACTGTGCC 857

```
|||||
Db 1 CAGCAATCATTAACAAGAGGCAATGATGGCGAGAGAGAGGTGGCCGACATGTC 60
Qy 858 ATGCTGCTCCAGACCCAAAGTGGAGCTCTCACTGCGGGACCTGCTACTTTGACAGC 917
Db 61 ATGCTGCTCCAGACCCAAAGTGGAGCTCTCACTGCGGGACCTGCTACTTTGACAGC 120
Qy 918 CATGCGGGCGGCATCTCTGGGCAATGATGGCTCTCTCAAGAGTGCAGAGAGCGCT 977
Db 121 CATGCGGGCGGCATCTCTGGGCAATGATGGCTCTCTCAAGAGTGCAGAGAGCGCT 180
Qy 978 GTACCCAGTGAAGTCAAGTCACTGCGAGACGCTCGGCTCAATGATCTTTGACAGAGAG 1037
Db 181 GTACCCAGTGAAGTCAAGTCACTGCGAGACGCTCGGCTCAATGATCTTTGACAGAGAG 240
Qy 1038 GACGTGGCGGCTGCTGTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1097
Db 241 GACGTGGCGGCTGCTGTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 299
Qy 1098 GAGATGGAGCTTCTCTCAGAGGCACTGACCCATCCGAGTGGACGTCGAGAGCGGCGC 1157
Db 300 GAGATGGAGCTTCTCTCAGAGGCACTGACCCATCCGAGTGGACGTCGAGAGCGGCGC 359
Qy 1158 CAATGGACGTCGAGGCTTCTCTGTCGTGACAGAGGAGGAGCTGCCACACCAAGAGCT 1217
Db 360 CAATGGACGTCGAGGCTTCTCTGTCGTGACAGAGGAGGAGCTGCCACACCAAGAGCT 419
Qy 1218 GCTGAGAGTCACTCTCCGTGTGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1277
Db 420 GCTGAGAGTCACTCTCCGTGTGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 479
Qy 1278 AGACTGTGGCGGCAAGAGCTGCGCGTGGACGCGATCGTGGAGAGCGGAGCACAGCTT 1337
Db 480 AGACTGTGGCGGCAAGAGCTGCGCGTGGACGCGATCGTGGAGAGCGGAGCACAGCTT 539
Qy 1338 GGGCGCGTGGCGGCAAGTCAAGCTTCTGCTATGATGAGACACACTCTGTGGGGATC 1397
Db 540 GGGCGCGTGGCGGCAAGTCAAGCTTCTGCTATGATGAGACACACTCTGTGGGGATC 599
Qy 1398 CTTGCTCTCCGGGAGTGGGTGTCGACAGCGGCCCACTGCTCCGGAGCGGAACCGGCT 1457
Db 600 CTTGCTCTCCGGGAGTGGGTGTCGACAGCGGCCCACTGCTCCGGAGCGGAACCGGCT 659
Qy 1458 CTTGCTCCGATGGAGAGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1515
Db 660 CTTGCTCCGATGGAGAGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 719
Qy 1516 CAGCT-GGGGGTGCAGGCTGTGTCCTACACAGGGGGCTATCTTCC-TTTGGAGACCCCA 1573
Db 720 CAGCTGGGGGGTGCAGGCTGTGTCCTACACAGGGGGCTATCTTCC-TTTGGGGAGCC 779
Qy 1574 ACACGAGAGAGAAACAGCAAGATATGCGCTGTGTCACCTCTCCAGTCCCTGCCCCCA 1633
Db 780 ACACGAGAGAGAAACAGCAAGATATGCGCTGTGTCACCTCTCTCACTCCCTGACG 839
Qy 1634 CAGATATCATCAGGCTGTGTCCTCCAGGTCGCGGCGCAGAGCC 1677
Db 840 AATTAATCA-----GCTGTGTGCTCTCCA-CTGCCGCGCAGGCC 877

RESULT 13
BX422288 961 bp mRNA linear EST 01-MAY-2004
LOCUS BX422288 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
DEFINITION BX422288 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
ACCESSION BX422288
VERSION BX422288.2 GI:46931436
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 961)
```

```
AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 13, 2003 this sequence version replaced gi:30655303.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremlieux, CP 5706 - 91057 Evry cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 9952.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CS0DM003BD05Q1&c=9952.f.

FEATURES
source
Location/Qualifiers
1..961
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM003YGI0"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match 31.9%; Score 753; DB 5; Length 961;
Best Local Similarity 97.4%; Pred. No. 2e-164;
Matches 817; Conservative 1; Mismatches 16; Indels 5; Gaps 5;

Qy 736 CCCCAGAGCCCTGCTCCCTCCATCTCTCAAGAGTCCACCTGAGCCAGAGGTCA 795
Db 78 CTTGAGGCTCCGCCCCCACTGTCGACCCCA-GGCTCCACCTGCGCCAGAGGTCA 136
Qy 796 GCCAGGAGATATTACAGAGGAGTGAATGCGCGAGAGAGAGGTGGCCGAGCTGTG 855
Db 137 GCCAGGAGATATTACAGAGGAGTGAATGCGCGAGAGAGAGGTGGCGGAGCTGTG 196
Qy 856 CCATGCTGTCGAGACCCAGAGTGGAGCTCTCACTGCGGGAGCCCTGACTTCTGACA 915
Db 197 CCATGCTGTCGAGACCCAGAGTGGAGCTCTCACTGCGGGAGCCCTGACTTCTGACA 256
Qy 916 GCCATCGGGGCGGATCTGGGCAATGATGAGCTGTTCTCTCAAGAGTGAACAGAGCCG 975
Db 257 GCCATCGGGGCGGATCTGGGCAATGATGAGCTGTTCTCTCAAGAGTGAACAGAGCCG 316
Qy 976 CTGTATCCAGTGCAGTCACTGCGGAGAGCTGCGCTCATGATGTTTGTGACAGAGAA 1015
Db 317 CTGTATCCAGTGCAGTCACTGCGGAGAGCTGCGCTCATGATGTTTGTGACAGAGAA 376
Qy 1036 GGGAGCGTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1095
Db 377 GGGAGCGTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 436
Qy 1096 GAGAGAGATGGGCTTCTCAGAGGCACTGACCCACTCCAGAGTGAAGTGCAGAGCGGGCC 1155
Db 437 GAGAGAGATGGGCTTCTCAGAGGCACTGACCCACTCCAGAGTGAAGTGCAGAGCGGGCC 496
Qy 1156 GCCAATGGCACTGCGGCTTCTTGTGTGAGAGAGGGAGGCTGCCACACCAAGAGG 1215
Db 497 GCCAATGGCACTGCGGCTTCTTGTGTGAGAGAGGGAGGCTGCCACACCAAGAGG 556
Qy 1216 CTGCTGAGAGTCACTCTCCGTGTGATGTCGCCAGAGAGCGGTTCTTGGCGCATCTGC 1275
Db 557 CTGCTGAGAGTCACTCTCCGTGTGATGTCGCCAGAGAGCGGTTCTTGGCGCATCTGC 616
Qy 1276 CAAGACTGTGGCGCAGAGAGCTGCCGCTGAGACCGCATCTGTGGAGGCGGAGCACAGC 1335
```

Accession	Sequence	Position
D6	617 CAGAGCTGTGGCCCGAGAAAGCTGTGCCCTCTGGACCGCATCTGTGGAGGCGGGGACACCAAGC	676
QY	1336 TTGGGCGCGGTGGCCCTGTGGCAAGTCAGGCTTTGGCTATGATGAGCAACACTTCTGTGGGGGA	1395
D6	677 TTGGGCGCGGTGGCCCTGTGGCAAGTCAGGCTTTGGCTATGATGAGCAACACTTCTGTGGGGGA	736
QY	1396 TCCCTGCTCTCCGGGAGCTGGGTGCTGACAGCCGCGCCACTGCTTCCCGGAGCGGAACCGG	1455
D6	737 T-CTGCTCTCCGGGAGCTGGGTGCTGACAGCCGCGCCACTGCTTCCCGGAGCGGA- CGG	794
QY	1456 GTCCTGTCCCGATGGCGAGGTGTTTGGGAGTCCGGAGCCGAGAGGCTCTCCGCAAGGCTG	1515
D6	795 GTCCTGTCCCGATGGCGAGGTGTTTGGGAGTCCGGAGCCGAGAGGCTCTCCGCAAGGCTG	854
QY	1516 CAGCTGGGGGTGCAGGCTGTGTGTCTACACAGGGGCTATCTTCCCTTTGGGAGCCCAA	1574
D6	855 CAGCTGGGGGT- CAGGCTGTGTGTCTA- CACGGGGCTATCTTCCCTTTGGGAGCCCAA	911

RESULT	14				
BX427780/C					
LOCUS	BX427780	674 bp	mRNA	linear	EST 03-MAY-2004
DEFINITION	BX427780 Homo sapiens FETAL LIVER Homo sapiens cDNA clone CSODM012Y015 3-PRIME, mRNA sequence.				
ACCESSION	BX427780				
VERSION	BX427780.2	GI:46959112			
KEYWORDS	EST.				
SOURCE	Homo sapiens	(human)			
ORGANISM	Homo sapiens				

REFERENCE	1 (bases 1 to 874)
AUTHORS	Li, W.B., Gruber, C., Jeasee, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	On May 15, 2003 this sequence version replaced gi:30770691.

Genoscope - Centre National de Séquençage
2 rue Gaston Crémieux, CP 5706 - 91057 Evry cedex - FRANCE
Email: segreg@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-c11g9 (dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 9952.f
For more information about this cluster, see
<http://www.genoscope.cns.fr/cna/s=CS0AMM012AH08NP1&c=9952.f>.

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM012Y015"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

```

Query Match 31.6%; Score 747; DB 5; Length 874;
 Best Local Similarity 95.6%; Pred. No. 4,9e-163;
 Matches 813; Conservative 8; Mismatches 20; Indels 9; Gaps 5;

QY 1465 CGATGCGCAGTGTTCGCGGTCGCGCCAGCGCTTCTCCCAAGGTCTTGCAAGCTGGG 1524
 DB 874 CGAATGGCAAKTKTGTGTSSTGAC---ATGCCCAACCTCTCCCAACGATGTGCAACT--GGG 819

OY		1525	GTCAGAGCTGTGGTCTACCAACGCGGGGCTATCTTCCCTTTGGGAAACCCCAACAGCAGAG	1584
Dd		818	GTGCAGAGCTGT- GTCCTACACGGGG---CTATCTTCCCTTGGAACCCCAACAGCAGAG	763
OY		1585	AACAGCAAACAATATGCGCCCTGGTCCACTCTCCAGTCCCCTGGCCCCCACAGAATCATC	1644
Dd		762	AACAGCAAACAATATTGCCCTGGTCCACTCTTCCAATCCCTGGCCCCCACAGAATCATC	703
OY		1645	CAGCCTGTGTGCTCCACAGCTGCGGCGAGGCCCTGGTGANTGGCAAGATCTGTACCCTG	1704
Dd		702	CAGCCTGTGTGCTCCACAGCTGCGGCGAGGCCCTGGTGANTGGCAAGATCTGTACCCTG	643
OY		1705	ACGGGCTGGGGCAACACGCACTATATGCGCAACAGCGCGGGTACTCTCAGAGGCTCGA	1764
Dd		642	ACGGGCTGGGGCAACACGCACTATATGCGCAACAGCGCGGGTACTCTCAGAGGCTCGA	583
OY		1765	GTCCCAATATTCAGCAATGATGTCTGCAATGGGGCTGACTTCTATGGAAAACGATCAAG	1824
Dd		582	GTCCCAATATTCAGCAATGATGTCTGCAATGGGGCTGACTTCTATGGAAAACGATCAAG	523
OY		1825	CCCAAGATGTCTGTGCTGGCTAACCCCGAGGGTGGCAATTGATGCTTGCACAGGCGACAGC	1884
Dd		522	CCCAAGATGTCTGTGCTGGCTAACCCCGAGGGTGGCAATTGATGCTTGCACAGGCGACAGC	463
OY		1885	GGTGTGCTCCCTTTGTGTGTAGAGAACAGCATCTTCTCGAACGCCAGTGGCGGCTGTGTGC	1944
Dd		462	GGTGTGCTCCCTTTGTGTGTAGAGAACAGCATCTTCTCGAACGCCAGTGGCGGCTGTGTGC	403
OY		1945	ATTGTGATTTGGGGCACTGGCTGTGCCCTTGGCCCAAGCCAGGCGCTTACACCAAAATGC	2004
Dd		402	ATTGTGATTTGGGGCACTGGCTGTGCCCTTGGCCCAAGCCAGGCGCTTACACCAAAATGC	343
OY		2005	AGTAGCTTCCGGGAGTGGATCTTCCAGGCGCATTAAGACTCACTCCGAAGCCAGCGGATG	2064
Dd		342	AGTAGCTTCCGGGAGTGGATCTTCCAGGCGCATTAAGACTCACTCCGAAGCCAGCGGATG	283
OY		2065	GTCACCAAGCTCTGACCGGTGGCTTTCGTGCGGCAAGCTTCAGAGGCCCGAGGTGATCCC	2124
Dd		282	GTCACCAAGCTCTGACCGGTGGCTTTCGTGCGGCAAGCTTCAGAGGCCCGAGGTGATCCC	224
OY		2125	GGTGTGGGANTCCACGCTGGGGCCGAGGATGGGACGTTTTCTTCTTGGGCGCGGTCCACA	2184
Dd		223	GGTGTGGGANTCCACGCTGGGGCCGAGGATGGGACGTTTTCTTCTTGGGCGCGGTCCACA	164
OY		2185	GGTCCAAAGACACCCCTCCCTCCAGGGGTCTCTTCCACAAGTGGGGGCCACTCAGAGCC	2244
Dd		163	GGTCCAAAGACACCCCTCCCTCCAGGGGTCTCTTCCACAAGTGGGGGCCACTCAGAGCC	104
OY		2245	CGAAGACCAACCACTCACCCCTCTGACCCCACTGAATAATATTTCTGTGCTGTGGAG	2304
Dd		103	CGAAGACCAACCAACCTCACCCCTCTGACCCCACTGAATAATATTTCTGTGCTGTGGAG	44
OY		2305	TCTGTGTTAG 2314	
Dd		43	ACCTGTCTNG 34	
RESULT_15				
Bt752774				
LOCUS				
DEFINITION			774 bp mRNA linear EST 25-SEP-2001	
ACCESSION			60302181.F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5192427 5'..	
VERSION			Bt752774	
KEYWORDS			Bt752774.1 GI:15744352	
SOURCE			EST.	
ORGANISM			Homo sapiens (human)	
REFERENCE			Homo sapiens	
AUTHORS			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
TITLE			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
JOURNAL			NIH-MGC http://mgc.ncl.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2005, 07:27:22 ; Search time 4014 Seconds

(without alignments)
3824.668 Million cell updates/sec

Title: US-10-054-498-1

Perfect score: 2363
Sequence: 1 tcgagccgcttcacggga.....taataagatggttgatt 2363

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapcut 1.0

Searched: 7316285 seqs, 321459403 residues

Total number of hits satisfying chosen parameters: 14632570

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10J_NEW_PUB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
25: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2363	100.0	2363	9	US-09-880-107-3690
2	2363	100.0	2363	11	US-09-968-007A-436
3	2363	100.0	2363	17	US-10-099-322-39
4	2363	100.0	2363	21	US-10-044-564-39
5	2363	100.0	2363	17	US-10-843-641A-6906
6	1593.6	67.4	1783	9	US-09-919-048-188
7	1593.6	67.4	1783	10	US-09-776-191-67

8	1593.6	67.4	1783	14	US-10-102-283-188	Sequence 188, App
9	1593.6	67.4	1783	14	US-10-073-060-1	Sequence 1, Appli
10	1593.6	67.4	1783	15	US-10-135-795-188	Sequence 188, App
11	1593.6	67.4	1783	15	US-10-205-823-177	Sequence 177, App
12	1593.6	67.4	1783	16	US-10-210-120-1	Sequence 1, Appli
13	1593.6	67.4	1783	17	US-10-295-027-1160	Sequence 1160, Ap
14	1593.6	67.4	1783	17	US-10-156-214A-34	Sequence 34, Appli
15	1593.6	67.4	1783	19	US-10-652-993-188	Sequence 188, App
16	1593.6	67.4	1783	21	US-10-956-157-792	Sequence 792, App
17	1593.6	67.4	1783	22	US-10-909-035-1	Sequence 1, Appli
18	1482	62.7	1482	9	US-09-822-830A-46	Sequence 46, Appli
19	1400	59.2	1400	21	US-10-956-157-6027	Sequence 6027, Ap
20	1269	53.7	1615	9	US-09-820-002-1	Sequence 1, Appli
21	1269	53.7	1615	15	US-10-274-031-1	Sequence 1, Appli
22	1254	53.1	1254	18	US-10-235-789-1	Sequence 1, Appli
23	1097.8	46.5	7033	19	US-10-678-816-5	Sequence 5, Appli
24	1097	46.4	12242	17	US-10-678-816-6	Sequence 6, Appli
25	858.4	36.3	2175	19	US-10-108-260A-559	Sequence 559, App
26	778.2	32.9	11924	19	US-10-678-816-7	Sequence 7, Appli
27	402.8	17.0	493	10	US-09-918-995-132762	Sequence 32762, A
28	380.2	16.1	497	19	US-10-250-889-38	Sequence 38, Appli
29	374.8	15.9	494	10	US-09-918-995-32775	Sequence 32775, A
30	325	13.8	21784	9	US-09-820-002-3	Sequence 3, Appli
31	325	13.8	21784	15	US-10-274-031-3	Sequence 3, Appli
32	321.8	13.6	2547	17	US-10-104-047-819	Sequence 819, App
33	279	11.8	279	18	US-10-424-599-133667	Sequence 133667, A
34	249.8	10.6	266	9	US-09-964-824A-462	Sequence 462, App
35	249.8	10.6	266	9	US-09-880-107-1812	Sequence 1812, Ap
36	249.8	10.6	266	21	US-10-843-641A-5765	Sequence 5765, App
37	192	8.1	601	9	US-09-820-002-15	Sequence 15, Appli
38	192	8.1	601	15	US-10-274-031-15	Sequence 15, Appli
39	191.4	8.1	568	16	US-10-029-386-2340	Sequence 2340, Ap
40	191	8.1	601	9	US-09-820-002-16	Sequence 16, Appli
41	191	8.1	601	15	US-10-274-031-16	Sequence 16, Appli
42	190	8.0	197	16	US-10-029-386-16040	Sequence 16040, A
43	185.6	7.9	379	9	US-09-960-352-7479	Sequence 7479, Ap
44	177.2	7.5	357	9	US-09-960-352-4196	Sequence 4196, Ap
45	169.4	7.2	601	9	US-09-820-002-14	Sequence 14, Appli

ALIGNMENTS

RESULT 1
US-09-880-107-3690
; Sequence 3690, Application US/09880107
; Patent No. US20020142981A1
GENERAL INFORMATION:
; APPLICANT: Horne, Darc T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherff, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3690
; LENGTH: 2363
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X07732
US-09-880-107-3690
Query Match 100.0%; Score 2363; DB 9; Length 2363;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGAGCCCCGCTTTCCAGGGACCCCTACCTGAGGGCCCAAGGTGAGGACAGCCCTGGCTTAC 60
Db 1 TCGAGCCCCGCTTTCCAGGGACCCCTACCTGAGGGCCCAAGGTGAGGACAGCCCTGGCTTAC 60
OY 61 AGGCCCAACAGCCAGCCCTCTGCTCAGAGCCGCCCGCTGCTGCGGGGCAACATGCTCC 120
Db 61 AGGCCCAACAGCCAGCCCTCTGCTCAGAGCCGCCCGCTGCTGCGGGGCAACATGCTCC 120
OY 121 TCGCCAGGCTGTGAGACCTGACCCGAGCCCGGCACTACCTCGAGGCTCCGCCCTCCACTGTC 180
Db 121 TCGCCAGGCTGTGAGACCTGACCCGAGCCCGGCACTACCTCGAGGCTCCGCCCTCCACTGTC 180
OY 131 TGGACCCCAAGGGGTAGAGGACAGAGGGCCCGAGCTGACAGTTCCAGCCCTGAGAGAGAGGG 240
Db 131 TGGACCCCAAGGGGTAGAGGACAGAGGGCCCGAGCTGACAGTTCCAGCCCTGAGAGAGAGGG 240
OY 241 TTCCTCATCCCCCAACCCAGCCCTATATGCCCACCTCTTAATAGAGGGTTCTTGAGGAC 300
Db 241 TTCCTCATCCCCCAACCCAGCCCTATATGCCCACCTCTTAATAGAGGGTTCTTGAGGAC 300
OY 301 TGAAGAGGGGACATATGACGTCTCCCAAGACCTAGGTGTTCTGTCTCTCTTCTTCTT 360
Db 301 TGAAGAGGGGACATATGACGTCTCCCAAGACCTAGGTGTTCTGTCTCTCTTCTTCTT 360
OY 361 CAGACTCAGCCGTTGAGACCCGAGCTCTTCTCCCAAGACCCAGAGATTTCCAGCCCTGAG 420
Db 361 CAGACTCAGCCGTTGAGACCCGAGCTCTTCTCCCAAGACCCAGAGATTTCCAGCCCTGAG 420
OY 421 GCCCTCCTCCTCATTAAGAGGATCTGAGCCGCCCAATTCCTCTTCCCAAGACTTA 480
Db 421 GCCCTCCTCCTCATTAAGAGGATCTGAGCCGCCCAATTCCTCTTCCCAAGACTTA 480
OY 481 TGAATTCAGGTCCTCAGCTGTCTCTCCCTCAACCCGGGATCTCAGTCCCTGTCTCAC 540
Db 481 TGAATTCAGGTCCTCAGCTGTCTCTCCCTCAACCCGGGATCTCAGTCCCTGTCTCAC 540
OY 541 CAGGCTCAGGCAATGAGGGGTCGCCCATCCCTGCAATTCAGAGGCTCCCGCGCTGCTGCA 600
Db 541 CAGGCTCAGGCAATGAGGGGTCGCCCATCCCTGCAATTCAGAGGCTCCCGCGCTGCTGCA 600
OY 601 GACACTGACCCCATCTTGAACCCAGCCCAATCTGCTCCGATCAGGCTGCTGCTG 660
Db 601 GACACTGACCCCATCTTGAACCCAGCCCAATCTGCTCCGATCAGGCTGCTGCTG 660
OY 661 CCAAGGCCCAAGTCCCTTACAGCTGCTGATGAGAGCCTGAGGAGCTGAGGGGCGCCAGACT 720
Db 661 CCAAGGCCCAAGTCCCTTACAGCTGCTGATGAGAGCCTGAGGAGCTGAGGGGCGCCAGACT 720
OY 721 GGGCTGGGCTGGGCTCCCGCAGGCTGCTCCGCTCATCTCTCAAGGTCACAGCC 780
Db 721 GGGCTGGGCTGGGCTCCCGCAGGCTGCTCCGCTCATCTCTCAAGGTCACAGCC 780
OY 781 TGGCCCAAGAGGTCAAGGAGGAAATTAACAAGAGGAGTGAACATGAGCGAGAGAG 840
Db 781 TGGCCCAAGAGGTCAAGGAGGAAATTAACAAGAGGAGTGAACATGAGCGAGAGAG 840
OY 841 GGTGGCCGAGCTGTGCTGCTGCTCAGACCCAGAGGTGAGAGCTTCACTGCGGGAGCC 900
Db 841 GGTGGCCGAGCTGTGCTGCTGCTCAGACCCAGAGGTGAGAGCTTCACTGCGGGAGCC 900
OY 901 CTGCTACTTCTGACAGGCAATGAGGGGCGGATCTGAGGCAATGAGGCTGCTCTCAAGG 960
Db 901 CTGCTACTTCTGACAGGCAATGAGGGGCGGATCTGAGGCAATGAGGCTGCTCTCAAGG 960
OY 961 AGTGACAGAGAGCCGCTGATCCAGTGAAGGTCAAGCTGCGAGAGCTCGGCTCATGATC 1020
Db 961 AGTGACAGAGAGCCGCTGATCCAGTGAAGGTCAAGCTGCGAGAGCTCGGCTCATGATC 1020
OY 1021 TTTTGAAGAAGAGGAG 1080
Db 1021 TTTTGAAGAAGAGGAG 1080

OY 1081 GCCGAGCTCAGCTGAGAGAGATGAGGCTTCTCTCAGGGGACATGACCCACTCCGAGCTGAG 1140
Db 1081 GCCGAGCTCAGCTGAGAGAGATGAGGCTTCTCTCAGGGGACATGACCCACTCCGAGCTGAG 1140
OY 1141 GTGCGAAGCGGAGGCGCCCAATGAGACGTGAGGCTTCTTCTGTGTGAGAGAGAGAGCTG 1200
Db 1141 GTGCGAAGCGGAGGCGCCCAATGAGACGTGAGGCTTCTTCTGTGTGAGAGAGAGAGCTG 1200
OY 1201 CCCCCAACCCAGAGGCTGTGAGAGTCACTCCGCTGTGATTTCCCGAGAGGCGCTTTC 1260
Db 1201 CCCCCAACCCAGAGGCTGTGAGAGTCACTCCGCTGTGATTTCCCGAGAGGCGCTTTC 1260
OY 1261 TTAGCCGCTCTCTCCAAAGCTGTGAGCGCAGAGAGCTCCCGTGAACCGCATGAGGAG 1320
Db 1261 TTAGCCGCTCTCTCCAAAGCTGTGAGCGCAGAGAGCTCCCGTGAACCGCATGAGGAG 1320
OY 1321 GAGCGGAGCACACGCTTGGGCGGATGAGCGGAGAGTCAAGCTTTCGCTATGATGAGACA 1380
Db 1321 GAGCGGAGCACACGCTTGGGCGGATGAGCGGAGAGTCAAGCTTTCGCTATGATGAGACA 1380
OY 1381 CACCTCTGTGAGGAGATCCCTGCTCTCGGAGGACCTGGGTGACAGCCGCCACTGCTTC 1440
Db 1381 CACCTCTGTGAGGAGATCCCTGCTCTCGGAGGACCTGGGTGACAGCCGCCACTGCTTC 1440
OY 1441 CCGAGAGCGAAGCCGAGTCTGTCCGATGAGCGAGTGTGTCGGGTGCGGTGCGGACAGG 1500
Db 1441 CCGAGAGCGAAGCCGAGTCTGTCCGATGAGCGAGTGTGTCGGGTGCGGTGCGGACAGG 1500
OY 1501 TCTCCCAACGAGTCTGACAGCTGAGGGGTGAGGCTGTGCTTACACAGGGGCTATCTTCC 1560
Db 1501 TCTCCCAACGAGTCTGACAGCTGAGGGGTGAGGCTGTGCTTACACAGGGGCTATCTTCC 1560
OY 1561 TTTTGGAGACCCCAACAG 1620
Db 1561 TTTTGGAGACCCCAACAG 1620
OY 1621 CCCCCTGCCCCACAGAAATATCATCAGCCTGTGTGCTCCAGCTGCGGACAGGCTGTG 1680
Db 1621 CCCCCTGCCCCACAGAAATATCATCAGCCTGTGTGCTCCAGCTGCGGACAGGCTGTG 1680
OY 1681 GTGATGAGCAAGATCTGTACCGGTGACGGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1740
Db 1681 GTGATGAGCAAGATCTGTACCGGTGACGGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1740
OY 1741 GCGGAGGAGCTCCAGAGAGGCTGAGGCTCCCAATATCAGAGAGAGAGAGAGAGAGAG 1800
Db 1741 GCGGAGGAGCTCCAGAGAGGCTGAGGCTCCCAATATCAGAGAGAGAGAGAGAGAGAG 1800
OY 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCTGACCCCGAGGAG 1860
Db 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCTGACCCCGAGGAG 1860
OY 1861 ATTATGCTTGCAGAGGCGAGACGCGGTGCTCTTGTGTGAGAGACAGCATCTCTCG 1920
Db 1861 ATTATGCTTGCAGAGGCGAGACGCGGTGCTCTTGTGTGAGAGACAGCATCTCTCG 1920
OY 1921 ACGCAGAGTGGAGGCTGTGTGAGATGTGAGGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 1921 ACGCAGAGTGGAGGCTGTGTGAGATGTGAGGAGAGAGAGAGAGAGAGAGAGAGAG 1980
OY 1981 AAGCAGAGGCTCTACCAAAATCAGTGAATCTTCGAGAGTGAATCTTCAGAGCCATAAG 2040
Db 1981 AAGCAGAGGCTCTACCAAAATCAGTGAATCTTCGAGAGTGAATCTTCAGAGCCATAAG 2040
OY 2041 ACTCATCTCCAGAGCAGCGGAGATGACCCAGCTTGAACCGGTGAGCTTCTGCTGCGCA 2100
Db 2041 ACTCATCTCCAGAGCAGCGGAGATGACCCAGCTTGAACCGGTGAGCTTCTGCTGCGCA 2100
OY 2101 GCTTTCAGAGGCGCAGAGTATCCCGGTGTGTGAGATTCACGCTGAGGCGAGAGAGAGAG 2160
Db 2101 GCTTTCAGAGGCGCAGAGTATCCCGGTGTGTGAGATTCACGCTGAGGCGAGAGAGAGAG 2160
OY 2161 TTTTCTTCTTGGGCGGATCCAGAGTCCAAAGACACCTCTCCAGAGGCTCTCTCTTC 2220

Db 1763 GACACTGACCCCATCTTGAAACCAAGCCCAATCTGGCTCCGATCAAGCGCTGCTGG 1704
Qy 661 CCAAGGCCCATGCTCTTACAGCTGCTGGATGAGGCTGGAGCTGGGGGGCCCAAGACT 720
Db 1703 CCAAGGCCCATGCTCTTACAGCTGCTGGATGAGGCTGGGGGGCCCAAGACT 1644
Qy 721 GGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 780
Db 1643 GGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGG 1584
Qy 781 TGGCCAGAGAGGTCAAGCCAGAGGATCATTAACAAGAGGAGTGAATGGCGAGAGAG 840
Db 1583 TGGCCAGAGAGGTCAAGCCAGAGGATCATTAACAAGAGGAGTGAATGGCGAGAGAG 1524
Qy 841 GGTGGCCGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Db 1523 GGTGGCCGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1464
Qy 901 CTGCTACTTCTGACAGCCATGGGGGGGGGATCTGGGGGGGATGGGGGGGATGGGGGGGAT 960
Db 1463 CTGCTACTTCTGACAGCCATGGGGGGGGGATCTGGGGGGGATGGGGGGGATGGGGGGGAT 1404
Qy 961 AGTGAACAGAGCCGCTGTATCCAGTGAAGTCACTGCGAGAGCTGCGCTCATGCTC 1020
Db 1403 AGTGAACAGAGCCGCTGTATCCAGTGAAGTCACTGCGAGAGCTGCGCTCATGCTC 1344
Qy 1021 TTTGACAAAGACGGAAGGAGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db 1343 TTTGACAAAGACGGAAGGAGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1284
Qy 1081 GCCGGAATCAAGTGGAGAGAGTGGGGCTTCTCAGGGGATGACCCATCTCGAGAGTGAAC 1140
Db 1283 GCCGGAATCAAGTGGAGAGAGTGGGGCTTCTCAGGGGATGACCCATCTCGAGAGTGAAC 1224
Qy 1141 GTGCAAGAGCGGCGGCGCCATGAGAGTGGAGAGTGGAGAGTGGAGAGTGGAGAGTGG 1200
Db 1223 GTGCAAGAGCGGCGGCGCCATGAGAGTGGAGAGTGGAGAGTGGAGAGTGGAGAGTGG 1164
Qy 1201 CCCCACACCCAGAGGCTGCTGAGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db 1163 CCCCACACCCAGAGGCTGCTGAGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1104
Qy 1261 TTGGCGGCGCATTTGCAAGAGTGTGGCGGAGAGAGTGGAGAGTGGAGAGTGGAGAGTGG 1320
Db 1103 TTGGCGGCGCATTTGCAAGAGTGTGGCGGAGAGAGTGGAGAGTGGAGAGTGGAGAGTGG 1044
Qy 1321 GGGCGGAGACACAGGCTTGGGGCGGAGTGGCGGAGAGTGAAGTCACTGCTGCTGCTGCTGCT 1380
Db 1043 GGGCGGAGACACAGGCTTGGGGCGGAGTGGCGGAGAGTGAAGTCACTGCTGCTGCTGCTGCT 984
Qy 1381 CACCTTGTGGGGAGTCCCTGCTCTCGGGGAGCTGGAGTGAAGCCGCGCCACCTGCTTC 1440
Db 983 CACCTTGTGGGGAGTCCCTGCTCTCGGGGAGCTGGAGTGAAGCCGCGCCACCTGCTTC 924
Qy 1441 CCGAGAGCGGAACCGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
Db 923 CCGAGAGCGGAACCGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 864
Qy 1501 TCTCCCAACAGGTCTGAGTGGGGGTGAGAGCTGTGCTCAACAGGGGGGCTATCTTCC 1560
Db 863 TCTCCCAACAGGTCTGAGTGGGGGTGAGAGCTGTGCTCAACAGGGGGGCTATCTTCC 804
Qy 1561 TTTGGGAGACCCCAACAG 1620
Db 803 TTTGGGAGACCCCAACAG 744
Qy 1621 CCCCAGCCCTCAAGAGATATATCAAGCTGTGAGTCCCAAGCTGCGGCGAGAGAGAGAGAG 1680
Db 743 CCCCAGCCCTCAAGAGATATATCAAGCTGTGAGTCCCAAGCTGCGGCGAGAGAGAGAGAG 684
Qy 1681 GTGATGAGCAAGATTTGATGAGTGAAGGAGTGGGGCAACAGCAGATATATGAGAGAG 1740
Db 683 GTGATGAGCAAGATTTGATGAGTGAAGGAGTGGGGCAACAGCAGATATATGAGAGAG 624

Qy 1741 GCGGGGATCTTCAGAGAGGCTTGAATCCCATTAATACAGATGATGTCTGCAATGGCGCT 1800
Db 623 GCGGGGATCTTCAGAGAGGCTTGAATCCCATTAATACAGATGATGTCTGCAATGGCGCT 564
Qy 1801 GACTTCTATGGAACCAATCAAGCCCAAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
Db 563 GACTTCTATGGAACCAATCAAGCCCAAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 504
Qy 1861 ATTATGCTGCTGCAAGGCGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
Db 503 ATTATGCTGCTGCAAGGCGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 444
Qy 1921 ACGCAGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
Db 443 ACGCAGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 384
Qy 1981 AAGCAGAGGCTCTACACCAAGTGAAGTCACTTCCGGAGTGGAGTCTTCAAGGCAATAAG 2040
Db 383 AAGCAGAGGCTCTACACCAAGTGAAGTCACTTCCGGAGTGGAGTCTTCAAGGCAATAAG 324
Qy 2041 ACTCACTCGAAGCGAGCGAGTGAAGCCAGCTTGAACCGTGGCTTCTGCTGCTGCTGCTG 2100
Db 323 ACTCACTCGAAGCGAGCGAGTGAAGCCAGCTTGAACCGTGGCTTCTGCTGCTGCTGCTG 264
Qy 2101 GCTTCAGAGGCGCGAGTGAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
Db 263 GCTTCAGAGGCGCGAGTGAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 204
Qy 2161 TTTTCTTGGGGCGGCTTCAAGAGTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 2220
Db 203 TTTTCTTGGGGCGGCTTCAAGAGTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 144
Qy 2221 CACAGTGGCGGCGCCACTCAAGCCCGAGAGAGCCCAACCTGACCTCTGAGAGTGGAGT 2280
Db 143 CACAGTGGCGGCGCCACTCAAGCCCGAGAGAGCCCAACCTGACCTCTGAGAGTGGAGT 84
Qy 2281 AAATATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340
Db 83 AAATATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24
Qy 2341 AAATATTAAGATGTTTGAAT 2363
Db 23 AAATATTAAGATGTTTGAAT 1

RESULT 4
US-10-044-564-39/c
; Sequence 39, Application US/10044564
; Publication No. US20040018196A1
; GENERAL INFORMATION:
; APPLICANT: Mezes et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-240
; CURRENT APPLICATION NUMBER: US/10/044,564
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/261,014
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,018
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/318,410
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/261,013
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,026
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/261,029
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/313,170
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39

LENGTH: 2363
TYPE: DNA
ORGANISM: Homo sapiens
US-10-044-564-39

Query Match 100.0%; Score 2363; DB 17; Length 2363;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TCGAGCCCGCTTTCCAGGAGCCCTTACCTGAGGGGCCCAAGGTGAGGCGAGCTGGCTTACG 60
2363 TCGAGCCCGCTTTCCAGGAGCCCTTACCTGAGGGGCCCAAGGTGAGGCGAGCTGGCTTACG 2304
61 AGGCCCCAGCCACGCGCTTGCCTCCAGAGCGCGCGCTGCTGCGGGGCCAATGCTTCC 120
2303 AGGCCCCAGCCACGCGCTTGCCTCCAGAGCGCGCGCTGCTGCGGGGCCAATGCTTCC 2244
121 TGCCAGGCGCTGAGAGCTGAGCCCGAGCCCGGCACTTCCAGAGCTTCGCGCCCACTGC 180
2243 TGCCAGGCGCTGAGAGCTGAGCCCGAGCCCGGCACTTCCAGAGCTTCGCGCCCACTGC 2184
181 TCGAGCCCGAGGTAAAGGACAAAGGGGCCCAAGACTCAAGTTCCAGCCCTGAGAGAGGGG 240
2183 TCGAGCCCGAGGTAAAGGACAAAGGGGCCCAAGACTCAAGTTCCAGCCCTGAGAGAGGGG 2124
241 TTCCCTCATCCCGCCAGCCAGCTTAATGCGCACTTAATGAGGGGTTCTGGGAGAC 300
2123 TTCCCTCATCCCGCCAGCCAGCTTAATGCGCACTTAATGAGGGGTTCTGGGAGAC 2064
301 TGAAGAGGGGACATGACGTCTCCCAAGAGCACTAGAGTTCTGTCTGTCTTCTT 360
2063 TGAAGAGGGGACATGACGTCTCCCAAGAGCACTAGAGTTCTGTCTGTCTTCTT 2004
361 CAGACTCAGCGGTTGAGACCCAGTCTTTCTCCCAAGACCAAGAGTTCCAGCCCTAG 420
2003 CAGACTCAGCGGTTGAGACCCAGTCTTTCTCCCAAGACCAAGAGTTCCAGCCCTAG 1944
421 GCCCTCTCTCTCTACTAGAGGAGTCCGAGCGGCCCAATTCTCTTCCCAAGACTTA 480
1943 GCCCTCTCTCTCTACTAGAGGAGTCCGAGCGGCCCAATTCTCTTCCCAAGACTTA 1884
481 TGATTTCAGGTCTCAGCTGTCTCTCTCCCAACCGGAGATCTCAGTCCCTGTCTCAC 540
1883 TGATTTCAGGTCTCAGCTGTCTCTCTCCCAACCGGAGATCTCAGTCCCTGTCTCAC 1824
541 CAGGCTCAGGCAATGGGGGTCCCATCTCTGCAAAATCCAGGGTCCCCCGCTGCTGCA 600
1823 CAGGCTCAGGCAATGGGGGTCCCATCTCTGCAAAATCCAGGGTCCCCCGCTGCTGCA 1764
601 GACACTGAGCCCATCTTGAACCCAGCCCAATCTGCTCCGATCAGGCGTGTCTGG 660
1763 GACACTGAGCCCATCTTGAACCCAGCCCAATCTGCTCCGATCAGGCGTGTCTGG 1704
661 CCAGAGCCCAATCTCTTACAGCTTGCCTGATGAGCGCTGAGACTGGGGGCGCCAGACT 720
1703 CCAGAGCCCAATCTCTTACAGCTTGCCTGATGAGCGCTGAGACTGGGGGCGCCAGACT 1644
721 GGGCTGGGCTGGGCTTCCCAAGGCGCTTCCCGTCAATCTCTCAAGGTCCCAACC 780
1643 GGGCTGGGCTGGGCTTCCCAAGGCGCTTCCCGTCAATCTCTCAAGGTCCCAACC 1584
781 TGGCCAGAGAGTCAAGGAGGAATCATTAACAAGGAGTGAATGAGGCGAGAAAGAG 840
1583 TGGCCAGAGAGTCAAGGAGGAATCATTAACAAGGAGTGAATGAGGCGAGAAAGAG 1524
841 GGTGGCCGGAAGTGTGCAATGCTGTCAAGCCCAAGAGTGGAGCTCTCACTGCGGGAGC 900
1523 GGTGGCCGGAAGTGTGCAATGCTGTCAAGCCCAAGAGTGGAGCTCTCACTGCGGGAGC 1464
901 CTGCTACTTCTGACAGGCAATGGGGGCGGATCTGGGGCAATTGGGCTGTCTCTCAGG 960
1463 CTGCTACTTCTGACAGGCAATGGGGGCGGATCTGGGGCAATTGGGCTGTCTCTCAGG 1404

961 AGTACGAGAGCGCGCTGTACCCAGTGCAGGTCAAGTCTGCGGAGCGCTCGGCTATGCTC 1020
1403 AGTACGAGAGCGCGCTGTACCCAGTGCAGGTCAAGTCTGCGGAGCGCTCGGCTATGCTC 1344
1021 TTTCACAAAGACGGAAGGAGCTGGCGGCTGTGTCTCTCTGCGCTCCAAAGCCAGGTA 1080
1343 TTTCACAAAGACGGAAGGAGCTGGCGGCTGTGTCTCTCTGCGCTCCAAAGCCAGGTA 1284
1081 GCGGAGCTCAGCTGCGAGAGATGGGGCTTCTCCAGGGCACTGACCACTCCGAGTGGAC 1140
1283 GCGGAGCTCAGCTGCGAGAGATGGGGCTTCTCCAGGGCACTGACCACTCCGAGTGGAC 1224
1141 GTGCGAAACGCGCGGCGCCCAATGCGACGTCGAGCTTCTTCTGTGTGAAGAGGAGCTG 1200
1223 GTGCGAAACGCGCGGCGCCCAATGCGACGTCGAGCTTCTTCTGTGTGAAGAGGAGCTG 1164
1201 CCCCAACACAGAGGCTGTGAGAGTATCTCCGTGTGTGATTGCCCAAGAGCGCTTTC 1260
1163 CCCCAACACAGAGGCTGTGAGAGTATCTCCGTGTGTGATTGCCCAAGAGCGCTTTC 1104
1261 TTGGCGGCAATCTGCGCAAGACTGTGGCGCGAGAGGCGCGGAGCCGATGTGGGA 1320
1103 TTGGCGGCAATCTGCGCAAGACTGTGGCGCGAGAGGCGCGGAGCCGATGTGGGA 1044
1321 GCGCGGAGACACAGCTTGGGCGGCTGGCCGTGGCAATGAGCTTTCGTATGATGAGACA 1380
1043 GCGCGGAGACACAGCTTGGGCGGCTGGCCGTGGCAATGAGCTTTCGTATGATGAGACA 984
1381 CACTCTGTGGGGGATCTCTGTCTCCGAGGAGCTGGGTTGCTGACAGCCCGCACTGCTTC 1440
983 CACTCTGTGGGGGATCTCTGTCTCCGAGGAGCTGGGTTGCTGACAGCCCGCACTGCTTC 924
1441 CCGGAGGGGAAACCGGGGTCTGTCCCGAATGGAGATGTTTGGCGGTGGCCGAGGCC 1500
923 CCGGAGGGGAAACCGGGGTCTGTCCCGAATGGAGATGTTTGGCGGTGGCCGAGGCC 864
1501 TCTCCCAACGATCTGACAGCTGGAGGATCAGGCTGTGTCTTACCAAGGGGCTATCTTCC 1560
863 TCTCCCAACGATCTGACAGCTGGAGGATCAGGCTGTGTCTTACCAAGGGGCTATCTTCC 804
1561 TTTTGGGAGCCCAACAGGAGAGAGACAGACATATTGCGCTGTGTCTTCACTTCCAGT 1620
803 TTTTGGGAGCCCAACAGGAGAGAGACAGACATATTGCGCTGTGTCTTCACTTCCAGT 744
1621 CCGCGCGCCCTCAAGAAATCATCAGCTGTGTGTGCTCCAGGCTGCGGCGAGGCGCTG 1680
743 CCGCGCGCCCTCAAGAAATCATCAGCTGTGTGTGCTCCAGGCTGCGGCGAGGCGCTG 684
1681 GTGATGGCAAGATCTGTACCGTGAACGGGCTGGGGCAACAGCAGTACTATGGCCAACAG 1740
683 GTGATGGCAAGATCTGTACCGTGAACGGGCTGGGGCAACAGCAGTACTATGGCCAACAG 624
1741 GCGGGGATCTCCAGAGGCTTCAAGTCCCAATATCAGCAATGATGTCTGCAATGGGCT 1800
623 GCGGGGATCTCCAGAGGCTTCAAGTCCCAATATCAGCAATGATGTCTGCAATGGGCT 564
1801 GACTTTATGGAACCAAGATCAAGCCCAAGATGTTCTGTGTGACTACCCGAGGGTGGC 1860
563 GACTTTATGGAACCAAGATCAAGCCCAAGATGTTCTGTGTGACTACCCGAGGGTGGC 504
1861 ATTGATGCTGCGCAGGCGAGCAGCGGTGTCTCTTGTGTGTGAAGCAGCATCTTGG 1920
503 ATTGATGCTGCGCAGGCGAGCAGCGGTGTCTCTTGTGTGTGAAGCAGCATCTTGG 444
1921 ACGCCAGCTTGGCGGCTGTGTGTGAAGTGTGGGCACTGGCTGTGCTTGGCCAG 1980
443 ACGCCAGCTTGGCGGCTGTGTGTGAAGTGTGGGCACTGGCTGTGCTTGGCCAG 384
1981 AAGCCAGCGCTTTCACCAAAAGTCAATGAGCTTCCGAGAGTGGAGTCTTCCAGGCAATAAG 2040
383 AAGCCAGCGCTTTCACCAAAAGTCAATGAGCTTCCGAGAGTGGAGTCTTCCAGGCAATAAG 324
2041 ACTCACTCCGAAGCAGCGGATGTGACCAAGCTTGACCGGCTTCTGCTGCGGCA 2100

1216 CTGCTGAGAGTCATCTCCGTGATGATTGCCCCAGAGGCGCTTTCTTGCCGCAATGTC 1275
 636 CTGCTGAGAGTCATCTCCGTGATGATTGCCCCAGAGGCGCTTTCTTGCCGCAATGTC 695
 1276 CAAGACTGTGGCCGAGAGAGTGGCCCGTGAACCGCATGTGGAGGCGGAGACACAGC 1335
 696 CAAGACTGTGGCCGAGAGAGTGGCCCGTGAACCGCATGTGGAGGCGGAGACACAGC 755
 1336 TTGGGCGGCTGGCCGTGCAAGTCAAGCTTGTGATGATGAGACACACTCTGTGGGGA 1395
 756 TTGGGCGGCTGGCCGTGCAAGTCAAGCTTGTGATGATGAGACACACTCTGTGGGGA 815
 1396 TCCCGCTCTCCGCGGAGCTGGGTGTGACGCGCGCCACTGTTCCCGGAGCGGAACGG 1455
 816 TCCCGCTCTCCGCGGAGCTGGGTGTGACGCGCGCCACTGTTCCCGGAGCGGAACGG 875
 1456 GTCCGTCTCCGCGGAGCTGGGTGTGACGCGCGCCACTGTTCCCGGAGCGGAACGG 1515
 876 GTCCGTCTCCGCGGAGCTGGGTGTGACGCGCGCCACTGTTCCCGGAGCGGAACGG 935
 1516 CAGCTGGGAGTGCAGGCTGTGGTCAACAGCGGAGCTATCTTCCCTTTGGGAGCCCAAC 1575
 936 CAGCTGGGAGTGCAGGCTGTGGTCAACAGCGGAGCTATCTTCCCTTTGGGAGCCCAAC 995
 1576 AGCGAGGAGAACAGAACGATATGCTGCTCCACTCTCCAGTCCCTCCCTCCCTCA 1635
 996 AGCGAGGAGAACAGAACGATATGCTGCTCCACTCTCCAGTCCCTCCCTCCCTCA 1055
 1636 GAATACATCCAGGCTGTGGTCAACAGCGGAGCTATCTTCCCTTTGGGAGCCCAAC 1695
 1056 GAATACATCCAGGCTGTGGTCAACAGCGGAGCTATCTTCCCTTTGGGAGCCCAAC 1115
 1696 TGTACCTGTGACGCGGTGGGAGAACAGCGAGTATATGAGCAACAGCGGAGTACTCAG 1755
 1116 TGTACCTGTGACGCGGTGGGAGAACAGCGAGTATATGAGCAACAGCGGAGTACTCAG 1175
 1756 GAGGCTGAGTCCCATATACAGCAATGATCTGCAATGCGCTGATCTTATGGAAC 1815
 1176 GAGGCTGAGTCCCATATACAGCAATGATCTGCAATGCGCTGATCTTATGGAAC 1235
 1816 CAGATCAAGCCCAAGATGTTCTGCTGCTGCTACCCGAGGCTGAGTATGATGCTGAG 1875
 1236 CAGATCAAGCCCAAGATGTTCTGCTGCTGCTACCCGAGGCTGAGTATGATGCTGAG 1295
 1876 GCGCAGAGCGGTGTCTCTTTGTGTGAGAGCAAGCATCTCTCGACGCCAGTGGCGG 1935
 1296 GCGCAGAGCGGTGTCTCTTTGTGTGAGAGCAAGCATCTCTCGACGCCAGTGGCGG 1355
 1936 CTGTGTGAGCATTTGTGAGTTGGGAGCACTGGCTGTGCTGCGCCAGAACCCAGGCGTCTAC 1995
 1356 CTGTGTGAGCATTTGTGAGTTGGGAGCACTGGCTGTGCTGCGCCAGAACCCAGGCGTCTAC 1415
 1996 ACCAAGTCAGTGACTTCGCGGAGTGTGATCTTCAAGGCGCATTAAGACATCTCCGAGCC 2055
 1416 ACCAAGTCAGTGACTTCGCGGAGTGTGATCTTCAAGGCGCATTAAGACATCTCCGAGCC 1475
 2056 AGCGGATGTGACCCAGCTGTGACCGGTGTGCTCTGCTGCGACGCTCCAGGAGCCCGA 2115
 1476 AGCGGATGTGACCCAGCTGTGACCGGTGTGCTCTGCTGCGACGCTCCAGGAGCCCGA 1535
 2116 GGTGATCTCCGCGGTGTGAGTCAAGCTGTGGGCGGAGATGGAACGTTTTTTCTTTGGGCG 2175
 1536 GGTGATCTCCGCGGTGTGAGTCAAGCTGTGGGCGGAGATGGAACGTTTTTTCTTTGGGCG 1595
 2176 CGGTTCACAGGTTCACAGACACCTCTCCAGGAGTCTCTCTTTCACAGTGGGCGGCGC 2235
 1596 CGGTTCACAGGTTCACAGACACCTCTCCAGGAGTCTCTCTTTCACAGTGGGCGGCGC 1655
 2236 ACTCAGCCCGGAGACACACCTCTCAGCTCTGAGACCCCAATGTAATATTTGTTCT 2295
 1656 ACTCAGCCCGGAGACACACCTCTCAGCTCTGAGACCCCAATGTAATATTTGTTCT 1715
 2296 GTCTGGAGCTCTGTCTAGTGTGCGGCTGATGAGGATGCTCTTAAATATAAAGATGG 2355

1716 GTCTGAGAGCTCTCTTCTAGAGTGGGCGGCTGATGATGAGTCTCTTAAATATAAAGATGG 1775
 2356 TTTTGATT 2363
 1776 TTTTGATT 1783

RESULT 7
 US-09-776-191-67
 ; Sequence 67, Application US/09776191
 ; Publication No. US20030119168A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Edwin L. Madison
 ; APPLICANT: Edgar O. Ong
 ; APPLICANT: Jiumn-Chern Yeh
 ; APPLICANT: Corvas International, Inc.
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
 ; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
 ; TITLE OF INVENTION: METHODS BASED THEREON
 ; FILE REFERENCE: 24745-1607
 ; CURRENT APPLICATION NUMBER: US/09/776,191
 ; PRIOR APPLICATION NUMBER: 60/213,124
 ; PRIOR FILING DATE: 2000-06-22
 ; PRIOR APPLICATION NUMBER: 60/234,840
 ; PRIOR FILING DATE: 2000-06-22
 ; PRIOR APPLICATION NUMBER: 60/179,982
 ; PRIOR FILING DATE: 2000-02-03
 ; PRIOR APPLICATION NUMBER: 60/183,542
 ; PRIOR FILING DATE: 2000-02-18
 ; PRIOR APPLICATION NUMBER: 09/657,968
 ; PRIOR FILING DATE: 2000-02-08
 ; NUMBER OF SEQ ID NOS: 72
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 67
 ; LENGTH: 1783
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (246)..(1499)
 ; OTHER INFORMATION: Nucleotide sequence encoding human hepsin
 ; PUBLICATION INFORMATION: GenBank M18930
 ; DATABASE ENTRY DATE: 1993-06-11
 ; US-09-776-191-67

Query Match 67.4%; Score 1593.6; DB 10; Length 1783;
 Best Local Similarity 99.1%; Pred. No. 0;
 Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

736 CCCCCAGGCGCTGCTCCCGCTGCATCTCCACAGGTCCACCCCTGGCCAGAGGTCA 795
 157 CTTGAGAGCTCCGCCCCCATCTGTGAGACCCA-GGATCCACCTTGCCAGAGGTCA 215
 796 GCCAGGGAATCATTAACAAGAGGAGTGAATGAGCGCAGAGAGAGGAGTGTG 855
 216 GCCAGGGAATCATTAACAAGAGGAGTGAATGAGCGCAGAGAGAGGAGTGTG 275
 856 CCATGCTGCTCCAGACCCCAAGGTGAGCTCTCACTGCGGGGACCTCTACTTCTACA 915
 276 CCAATGCTGCTCCAGACCCCAAGGTGAGCTCTCACTGCGGGGACCTCTACTTCTACA 335
 916 GCCATGCGGGCGGATCTGAGGCAATGTTGGCTGTTCTCTCAGAGAGTGAACAGAGCCG 975
 336 GCCATGCGGGCGGATCTGAGGCAATGTTGGCTGTTCTCTCAGAGAGTGAACAGAGCCG 395
 976 CTGTACCAAGTGAAGTCAAGCTCTGAGAGCGCTTGCTCATGATGTTTGAACAAGCGAA 1035
 396 CTGTACCAAGTGAAGTCAAGCTCTGAGAGCGCTTGCTCATGATGTTTGAACAAGCGAA 455
 1036 GGAGCTGTGGGCTGCTGTCTCTCTGCGGCTCCACCGCAGGAGTGAAGCGGATCAAGTGC 1095

Db	456	GGAGAGTGGCGGCTCTGTGCTTCTCGGCTCAACGCGAGGTAGCGGACTCAGCTTCG	515
Qy	1096	GAGGAGATGGGCTTCTCTCAGGCACTGACCCCATCTCCAGCTGGAAGTGCAGACGGCGGGC	1155
Db	516	GAGGAGATGGGCTTCTCTCAGGCACTGACCCCATCTCCAGAGCTGGAAGTGCAGACGGCGGGC	575
Qy	1156	GCCATATGGCACTCGGGCTCTTCTCTGTGTGACAGAGGGGAGGCTCCCAACCCAGAG	1215
Db	576	GCCATATGGCACTCGGGCTCTTCTCTGTGTGACAGAGGGGAGGCTCCCAACCCAGAG	635
Qy	1216	CTGCTGAGAGTATCTCCGTGTGTATTTGCCAGAGGCGCTTTCTTGAGCGCAATCTGC	1275
Db	636	CTGCTGAGAGTATCTCCGTGTGTATTTGCCAGAGGCGCTTTCTTGAGCGCAATCTGC	695
Qy	1276	CAGAAGTGTGCGCGAGGAGAGCTGCGCCGTGACCGCATCTGTGGAGGCGGAGCACAGC	1335
Db	696	CAGAAGTGTGCGCGAGGAGAGCTGCGCCGTGACCGCATCTGTGGAGGCGGAGCACAGC	755
Qy	1336	TTGGGCGCGTGGCCGTGGCAATCAGCTTCCTGTATGATGAGACACACTCTGTGGGGGA	1395
Db	756	TTGGGCGCGTGGCCGTGGCAATCAGCTTCCTGTATGATGAGACACACTCTGTGGGGGA	815
Qy	1396	TCCCGCTCTCCGGGGAGCTGGGATGCTGACAGCGCGCCACTGCTTCCCGAGGAGAACGG	1455
Db	816	TCCCGCTCTCCGGGGAGCTGGGATGCTGACAGCGCGCCACTGCTTCCCGAGGAGAACGG	875
Qy	1456	GTCCTGTCCCGATGCGAGTGTTCGCGGTGCGCTGAGGCGCAGGCTCTCCCAAGGCTG	1515
Db	876	GTCCTGTCCCGATGCGAGTGTTCGCGGTGCGCTGAGGCGCAGGCTCTCCCAAGGCTG	935
Qy	1516	CAGCTGGGAGTGCAGGCTGTGGTCTAACCAAGGGGCTATCTTCCCTTGGGAGCCCAAC	1575
Db	936	CAGCTGGGAGTGCAGGCTGTGGTCTAACCAAGGGGCTATCTTCCCTTGGGAGCCCAAC	995
Qy	1576	AGCGAGGAGAACAGCAAGATTTGCGCTGATCAACAGGGGCTATCTTCCCTTGGGAGCCCAAC	1635
Db	996	AGCGAGGAGAACAGCAAGATTTGCGCTGATCAACAGGGGCTATCTTCCCTTGGGAGCCCAAC	1055
Qy	1636	GAATATCATTCAGGCTGTGTGCTCTCCAGACTGCGGCGCAGAGCCCTGTGATATGCAAGATC	1685
Db	1056	GAATATCATTCAGGCTGTGTGCTCTCCAGACTGCGGCGCAGAGCCCTGTGATATGCAAGATC	1115
Qy	1686	TGTACCGTGAAGCGGCTGGGGCAACAGCAATCTATGCGCAACAGAGCGGGGATCTCCAG	1755
Db	1116	TGTACCGTGAAGCGGCTGGGGCAACAGCAATCTATGCGCAACAGAGCGGGGATCTCCAG	1175
Qy	1756	GAGGCTCGAGTCCCATATATCAGCAATGATGTCTGCAATGGCGCTGACTTCTATGGAAC	1815
Db	1176	GAGGCTCGAGTCCCATATATCAGCAATGATGTCTGCAATGGCGCTGACTTCTATGGAAC	1235
Qy	1816	CAGATCAAGCCCAAGATGTTCTGTACTGAGCTACCCCGAGGGGTGCAATTGATGCTGCAG	1875
Db	1236	CAGATCAAGCCCAAGATGTTCTGTACTGAGCTACCCCGAGGGGTGCAATTGATGCTGCAG	1295
Qy	1876	GCGGACAGCGGTGATCCCTTTGTGTGATGAGACAGCAATCTCTCGAGCGCACGTTGGGG	1935
Db	1296	GCGGACAGCGGTGATCCCTTTGTGTGATGAGACAGCAATCTCTCGAGCGCACGTTGGGG	1355
Qy	1936	CTGTGTGGCATTTGTAGTTGGGGCACTGGCTGTGCTGCGCCAGAAAGCAGGCGTCTAC	1995
Db	1356	CTGTGTGGCATTTGTAGTTGGGGCACTGGCTGTGCTGCGCCAGAAAGCAGGCGTCTAC	1415
Qy	1996	ACCAAAAGTCAGTCACTTCGCGGAGTGGATCTTCAGAGGCATPAAAGTCATCTCCGAAGCC	2055
Db	1416	ACCAAAAGTCAGTCACTTCGCGGAGTGGATCTTCAGAGGCATPAAAGTCATCTCCGAAGCC	1475
Qy	2056	AGCGGCATGTGACCCAGCTGACCGGATGCTTCTGCTGCGACGCTCCAGAGGCCGGA	2115
Db	1476	AGCGGCATGTGACCCAGCTGACCGGATGCTTCTGCTGCGACGCTCCAGAGGCCGGA	1535
Qy	2116	GGTGATCCCGGTGTGGGATTCACGCTGGCCGAGGATGGAGCGTTTTTCTTCTTGGGCC	2175

Db	1536	GGTATATCCCGGTGGGTGGATATCCAGCTGTGGGCCGAGATGGGAAAGTTTTTCTTCTTGGGCC	1599
Qy	2176	CGGTTCACAGGTCCAAAGGACACCTTCCCTCCAGGGTCTCTCTTCCACAGTGGCGGGCCC	2233
Db	1596	CGGTTCACAGGTCCAAAGGACACCTTCCCTCCAGGGTCTCTCTTCCACAGTGGCGGGCCC	1655
Qy	2236	ACTAGAGCCCGGAGACACCCAAAGCTCACCCCTCGAGCCCGCATGTAATATGTTCTGCT	2299
Db	1656	ACTAGAGCCCGGAGACACCCAAAGCTCACCCCTCGAGCCCGCATGTAATATGTTCTGCT	1715
Qy	2296	GTCTGGGACTCTGTCTTAGTGTGCCCTGATGATGGATGCTCTTTAAATAAATAAGATGG	2355
Db	1716	GTCTGGGACTCTGTCTTAGTGTGCCCTGATGATGGATGCTCTTTAAATAAATAAGATGG	1779
Qy	2356	TTTTTGATT 2363	
Db	1776	TTTTTGATT 1783	
RESULT 8			
US-10-102-283-188			
; Sequence 188, Application US/10102283			
; Publication No. US20030027181A1			
GENERAL INFORMATION:			
; APPLICANT: O'Brien, Timothy J.			
; APPLICANT: Cannon, Martin J.			
; APPLICANT: Santin, Alessandro			
; TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer			
; FILE REFERENCE: D6223CIP/A/D/CIP2			
; CURRENT APPLICATION NUMBER: US/10/102,283			
; PRIOR FILING DATE: 2002-03-20			
; PRIOR APPLICATION NUMBER: 09/919,048			
; PRIOR FILING DATE: 2001-07-30			
; NUMBER OF SEQ ID NOS: 190			
; SEQ ID NO 188			
; LENGTH: 1783			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; OTHER INFORMATION: full length cDNA of hepsin			
US-10-102-283-188			
Query Match 67.4%; Score 1593.6; DB 14; Length 1783;			
Best Local Similarity 99.1%; Pred. No. 0;			
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;			
Qy	736	CCCCAGGCGCTGCGTCCCGTCCATCTCTTCACAGGTCCCAAGTCCCAAGCTGGCCAGAGGTCA	795
Db	157	CTTGAGGCTCGGCCCACTGCTGCAGACCCCA-GGGTCCACCTTGGCCAGAGGTCA	215
Qy	796	GCCAGGGAGTATTAAACAAGAGGAGTGCACATGCGCGAGAGAGAGGGTGGCCGAGCTGTG	855
Db	216	GCCAGGGAGTATTAAACAAGAGGAGTGCACATGCGCGAGAGAGAGGGTGGCCGAGCTGTG	275
Qy	856	CCATGCTGCTCCAGACCCCAAGGTGGAGCTTCACTGCGGGGACCCCTGCTACTTCTGACA	915
Db	276	CCATGCTGCTCCAGACCCCAAGGTGGAGCTTCACTGCGGGGACCCCTGCTACTTCTGACA	335
Qy	916	GCCATCGGGGGGGGATCTGGGCGCATTTGTGGCTGTTCTCCACAGAGTGCACAGAGCGG	975
Db	336	GCCATCGGGGGGGGATCTGGGCGCATTTGTGGCTGTTCTCCACAGAGTGCACAGAGCGG	395
Qy	976	CTGATACCAAGTGCAGGTACGCTCTGCGGACGCTCGGCTCAATGATGTCCTTGAACAAGCGAA	1039
Db	396	CTGATACCAAGTGCAGGTACGCTCTGCGGACGCTCGGCTCAATGATGTCCTTGAACAAGCGAA	455
Qy	1036	GGAGCGTGGCGGCTGTGTCTCTCGCGCTCCAAAGCCAGGGTACCGGACTCACTGC	1099
Db	456	GGAGCGTGGCGGCTGTGTCTCTCGCGCTCCAAAGCCAGGGTACCGGACTCACTGC	515
Qy	1096	GAGAGAGATGGGCTTCTTCAAGGGACATGACCACTCCAGCTGAGACGTGCCAACGGCGGGC	1155
Db	516	GAGAGAGATGGGCTTCTTCAAGGGACATGACCACTCCAGCTGAGACGTGCCAACGGCGGGC	575


```
Db 696 CAAGACTGTGGCCCGAGAGAACTGCCCCCTGGACCCGACCTGTGGAGGCGCGGAGACCCAGC 755
Qy 1336 TTGGGCGCGGTGACCGGTGGCAAGTCAGCCTTTCGTATGATGAGACACACTCTGTGTGGGGA 1395
Db 756 TTGGGCGCGGTGACCGGTGGCAAGTCAGCCTTTCGTATGATGAGACACACTCTGTGTGGGGA 815
Qy 1396 TCCCTGCTCTCCGGGGAATGGGTGTGTGACCGCGCCCACTGCTTCCCGGAGACGGAACCGG 1455
Db 816 TCCCTGCTCTCCGGGGAATGGGTGTGTGACCGCGCCCACTGCTTCCCGGAGACGGAACCGG 875
Qy 1456 GTCCCTGCTCCCGATGAGCGATGTTTGGCCGGTGGCCGATGAGCCGCTCTCCCAAGCTGTG 1515
Db 876 GTCCCTGCTCCCGATGAGCGATGTTTGGCCGGTGGCCGATGAGCCGCTCTCCCAAGCTGTG 935
Qy 1516 CAGCTGGGGGTGACAGCTGTGTGTACACAGGGGGCTATCTTCCCTTTGGGAGACCCAGC 1575
Db 936 CAGCTGGGGGTGACAGCTGTGTGTGTACACAGGGGGCTATCTTCCCTTTGGGAGACCCAGC 995
Qy 1576 AGCGAGGAGAACAGCAAGATATGTCCTGGTCCACTCTCAGTCCCTGCCCCCTCA 1635
Db 996 AGCGAGGAGAACAGCAAGATATGTCCTGGTCCACTCTCAGTCCCTGCCCCCTCA 1055
Qy 1636 GAATACATCCAGCCTGTGTGCTCCCAAGCTGCGGCGCAAGCCCTGTGTGATGAGCAAGATC 1695
Db 1056 GAATACATCCAGCCTGTGTGCTCCCAAGCTGCGGCGCAAGCCCTGTGTGATGAGCAAGATC 1115
Qy 1696 TGTACCGTGTGACGGGTGGGAGACAGCAAGTATATGAGCAAGGCGGGGTACTCCAG 1755
Db 1116 TGTACCGTGTGACGGGTGGGAGACAGCAAGTATATGAGCAAGGCGGGGTACTCCAG 1175
Qy 1756 GAGGCTGAGTCCCATATACAGCAATGATGTGCAATGAGCGCTGATCTTATGGAAC 1815
Db 1176 GAGGCTGAGTCCCATATACAGCAATGATGTGCAATGAGCGCTGATCTTATGGAAC 1235
Qy 1816 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGATGAGTATGATGCTGCCAG 1875
Db 1236 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGATGAGTATGATGCTGCCAG 1295
Qy 1876 GCGCAGACGGGTGTCTCTTTGTGTGTGAGACAGCATCTCTCGAGCGCCAGTTGGCGG 1935
Db 1296 GCGCAGACGGGTGTCTCTTTGTGTGTGAGACAGCATCTCTCGAGCGCCAGTTGGCGG 1355
Qy 1936 CTGTGTGCTATTTGAGTGTGGGGCACTGGCTGTGCTGCGCCGAGGAGCGAGGCTTAC 1995
Db 1356 CTGTGTGCTATTTGAGTGTGGGGCACTGGCTGTGCTGCGCCGAGGAGCGAGGCTTAC 1415
Qy 1996 ACCAAAGTCAGTGACTTCCGGGAGTGAATCTTCCAGGCGCATTAAGACTCACTCCGAAGCC 2055
Db 1416 ACCAAAGTCAGTGACTTCCGGGAGTGAATCTTCCAGGCGCATTAAGACTCACTCCGAAGCC 1475
Qy 2056 AGCGGCAATGTGACCCAGACTCTGACCGGTGGCTTCTCGCTGGCGACGCTCCAGGGCCGA 2115
Db 1476 AGCGGCAATGTGACCCAGACTCTGACCGGTGGCTTCTCGCTGGCGACGCTCCAGGGCCGA 1535
Qy 2116 GGTGATCCCGGTGGTGGATCCAGCTGGGCGAGATGGAGAGCTTTTCTTCTTGGGCC 2175
Db 1536 GGTGATCCCGGTGGTGGATCCAGCTGGGCGAGATGGAGAGCTTTTCTTCTTGGGCC 1595
Qy 2176 CGGTTCACAGGTCCAGAGACACCTCTCCAGGGTCTCTTCTTCCAGTGGCGGGCC 2235
Db 1596 CGGTTCACAGGTCCAGAGACACCTCTCCAGGGTCTCTTCTTCCAGTGGCGGGCC 1655
Qy 2236 ACTCAGCCCGGAGACACCCAACTCAGCCCTCTGACCCCAATGTAATATTTGTCT 2295
Db 1656 ACTCAGCCCGGAGACACCCAACTCAGCCCTCTGACCCCAATGTAATATTTGTCT 1715
Qy 2296 GTCTGGGACTGCTGTCTAGGTGCTCTGATGATGGAGTCTTTAAATATTAAGATGG 2355
Db 1716 GTCTGGGACTGCTGTCTAGGTGCTCTGATGATGGAGTCTTTAAATATTAAGATGG 1775
Qy 2356 TTTTGATT 2363
|||||
```

```
Db 1776 TTTTGATT 1783
RESULT 10
US-10-135-795-188
; Sequence 188; Application US/10135795
; Publication No. US20030077618A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Cannon, Martin J.
; APPLICANT: Santlin, Alessandro
; TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer
; FILE REFERENCE: D6223CIP/A/D/CIP3
; CURRENT APPLICATION NUMBER: US/10/135,795
; PRIOR FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2002-03-20
; NUMBER OF SEQ ID NOS: 191
; SEQ ID NO 188
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: full length cDNA of hepsin
US-10-135-795-188
Query March 67.4%; Score 1593.6; DB 14; Length 1783;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
Qy 736 CCCCCAGGCGCTGCTCCCGTCCATCTCTCAAGAGTCCACCTGAGCCAGAGAGTCA 795
Db 157 CTTGAGAGCTTCGCGCCCACTGCTGAGACCCCA-GGGTCCACCTTGCCAGAGAGTCA 215
Qy 796 GCCAGGGAATCATTAACAAGAGCAGTACATGACGCGCAGAAAGAGGTTGCCGACTGTG 855
Db 216 GCCAGGGAATCATTAACAAGAGCAGTACATGACGCGCAGAAAGAGGTTGCCGACTGTG 275
Qy 856 CCATGCTGCTCCAGACCCCAAGTGGGAGCTCTCACTGCGGGAGCCGCTTACTTGTACA 915
Db 276 CCATGCTGCTCCAGACCCCAAGTGGGAGCTCTCACTGCGGGAGCCGCTTACTTGTACA 335
Qy 916 GCCATCGGGCGGAGATCTGGGCAATGTGGCTGTTCTCTCAGAGTGAACAGAGCCG 975
Db 336 GCCATCGGGCGGAGATCTGGGCAATGTGGCTGTTCTCTCAGAGTGAACAGAGCCG 395
Qy 976 CTGTACCCAGTGAAGTACGCTCTGCGGACGCTCGGCTCATGCTCTTTGAACAAGCGAA 1035
Db 396 CTGTACCCAGTGAAGTACGCTCTGCGGACGCTCGGCTCATGCTCTTTGAACAAGCGAA 455
Qy 1036 GGGACGTGGGGCTGTGCTGCTCTCTGCGCTCCAAAGCGGAGTGGAGCTCAGCTGC 1095
Db 456 GGGACGTGGGGCTGTGCTGCTCTCTGCGCTCCAAAGCGGAGTGGAGCTCAGCTGC 515
Qy 1096 GAGGAGATGGGCTTCTTCAAGGCACTGACCCACTCCAGCTGAGCTGAGCGGCGAGC 1155
Db 516 GAGGAGATGGGCTTCTTCAAGGCACTGACCCACTCCAGCTGAGCTGAGCGGCGAGC 575
Qy 1156 GCCAATGGCAAGTGGGCTTCTTCTGTGTGAGAGAGGAGGCTGCCCAACCCAGAGG 1215
Db 576 GCCAATGGCAAGTGGGCTTCTTCTGTGTGAGAGAGGAGGCTGCCCAACCCAGAGG 635
Qy 1216 CTGCTGGAAGTCAATCTCCGTGTGATGATGAGCCAGAGCGGTTCTTGGCGCATGTGC 1275
Db 636 CTGCTGGAAGTCAATCTCCGTGTGATGATGAGCCAGAGCGGTTCTTGGCGCATGTGC 695
Qy 1276 CAAGACTGTGAGCGCAGAGAGCTGCGCTGGAACGACTCTGTGGAGGCGCGGACACAGC 1335
Db 696 CAAGACTGTGAGCGCAGAGAGCTGCGCTGGAACGACTCTGTGGAGGCGCGGACACAGC 755
Qy 1336 TTGGGCGCGGTGGCGGTGGCAAGTCAAGCTTTCGTATGATGAGACACCTCTGTGGGGGA 1395
Db 756 TTGGGCGCGGTGGCGGTGGCAAGTCAAGCTTTCGTATGATGAGACACCTCTGTGGGGGA 815
```

1396 TCCCTGCTCCGGGAGTGGGTGCTGACAGCCGCCCACTGTTCCCGAGCGGAACCGG 1455
Db TCCCTGCTCCGGGAGTGGGTGCTGACAGCCGCCCACTGTTCCCGAGCGGAACCGG 875
1456 GTCCGTGCTCCGATGGAGTGTGTTCCGGGTGCGGTGGCCAGAGGCTCTCCCAAGGTCTG 1515
Db GTCCGTGCTCCGATGGAGTGTGTTCCGGGTGCGGTGGCCAGAGGCTCTCCCAAGGTCTG 935
1516 CAGCTGGGGGTGACAGGCTGTGTCTACACCGGGGGCTATCTTCCCTTTCGGGACCCCAAC 1575
Db CAGCTGGGGGTGACAGGCTGTGTCTACACCGGGGGCTATCTTCCCTTTCGGGACCCCAAC 995
1576 AGCGAGGAGAACAGACGATATGCTGCTGCTGACCTCTGACAGTCCCTGCTCCCTCA 1635
Db AGCGAGGAGAACAGACGATATGCTGCTGCTGACCTCTGACAGTCCCTGCTCCCTCA 1055
1636 GAATACATCCAGCTGTGTGCTCTCCAGCTGCGGACAGGCTGCTGGATGGCAAGATC 1695
Db GAATACATCCAGCTGTGTGCTCTCCAGCTGCGGACAGGCTGCTGGATGGCAAGATC 1115
1696 TGTACCGTACCGGCTGGGGCAACGAGATCTATGCGCAACAGGCGGGGTACTCCAG 1755
Db TGTACCGTACCGGCTGGGGCAACGAGATCTATGCGCAACAGGCGGGGTACTCCAG 1175
1756 GAGGCTGAGTCCCATATATACGACATATGTCTGCAATGGCGCTGACTTCTATGGAAC 1815
Db GAGGCTGAGTCCCATATATACGACATATGTCTGCAATGGCGCTGACTTCTATGGAAC 1235
1816 CAGATCAAGCCCAAGATGTTCTGTGCTGCTGCTACCCGAGGGGTGCTGATGGCTGGCAG 1875
Db CAGATCAAGCCCAAGATGTTCTGTGCTGCTGCTACCCGAGGGGTGCTGATGGCTGGCAG 1295
1876 GCGGACAGCGGTGTCTCTTGTGTGTGAGACAGCATCTCTCGACGCCAGTGGCGG 1935
Db GCGGACAGCGGTGTCTCTTGTGTGTGAGACAGCATCTCTCGACGCCAGTGGCGG 1355
1936 CTGTGTGCTGATTTGATGTTGGGGCACTGCTGTGCTGCTGCGCCAGAACCCGAGCTTAC 1995
Db CTGTGTGCTGATTTGATGTTGGGGCACTGCTGTGCTGCTGCGCCAGAACCCGAGCTTAC 1415
1996 ACCAAGTCAGTGAATTCGGGAGTGTGATCTTCCAGGCTATTAAGATCATCTCCGAGACC 2055
Db ACCAAGTCAGTGAATTCGGGAGTGTGATCTTCCAGGCTATTAAGATCATCTCCGAGACC 1475
2056 AGCGGATGTGACCCAGCTGACCGGTGTCTCTGCTGCGACGCTCCAGGCGCCGA 2115
Db AGCGGATGTGACCCAGCTGACCGGTGTCTCTGCTGCGACGCTCCAGGCGCCGA 1535
2116 GGTGATCCCGGTGTGGGATCCAGCTGGGCGGAGATGGAGCGTTTCTTCTTGGGCGC 2175
Db GGTGATCCCGGTGTGGGATCCAGCTGGGCGGAGATGGAGCGTTTCTTCTTGGGCGC 1595
2176 CGGTCCACAGGTCCAAAGGACACCTCCCTCCAGAGGTCTCTTCCCAAGTGGGGGGCC 2235
Db CGGTCCACAGGTCCAAAGGACACCTCCCTCCAGAGGTCTCTTCCCAAGTGGGGGGCC 1655
2236 ACTGAGCCCGAGACCAACCACTCACTCTGACCCCGCATGTAATATTTGTTCTGCT 2295
Db ACTGAGCCCGAGACCAACCACTCACTCTGACCCCGCATGTAATATTTGTTCTGCT 1715
2296 GTCTGGGACTCTGTCTGATGTTGGGCTGATGATGGATGCTCTTAAATTAAGATGG 2355
Db GTCTGGGACTCTGTCTGATGTTGGGCTGATGATGGATGCTCTTAAATTAAGATGG 1775
2356 TTTTGATT 2363
Db TTTTGATT 1783

RESULT 11
US-10-205-823-177
; Sequence 177, Application US/10205823

Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Monney, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Duclun
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205, 823
; PRIOR FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307, 982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314, 356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325, 020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341, 746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362, 158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 177
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-177
Query Match 67.4%; Score 1593.6; DB 15; Length 1783;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
Db 736 CCCCCAGGCCCCGCTCCCGTCCATCTCCCTCAAGGTCGCCAGCTGGCCCGAGAGTCA 795
Db 157 CTTGAGGCTTCGCCCCCACTGCTGACCCCA-GGTTCCACCTGGCCCGAGAGTCA 215
Db 796 GCCAGGATCATTTAACAAGAGGAGTACATGCGGAGAGAGAGGAGTGGCGGACTGTG 855
Db 216 GCCAGGATCATTTAACAAGAGGAGTACATGCGGAGAGAGAGGAGTGGCGGACTGTG 275
Db 856 CCATGCTGCTCAGACCCCAAGTGGCACTCTCACTCGGGGAGACCTGTCTACTGACA 915
Db 276 CCATGCTGCTCAGACCCCAAGTGGCACTCTCACTCGGGGAGACCTGTCTACTGACA 335
Db 916 GCCATCGGGGGGAGATCTGGGCAATGTGGCTGTCTCTCAGAGATCAAGAGCGG 975
Db 336 GCCATCGGGGGGAGATCTGGGCAATGTGGCTGTCTCTCAGAGATCAAGAGCGG 395
Db 976 CTGTACCAAGTGCAGTCACTGTCGCGAGCTCGGCTCAATGCTCTTGTGACAGAGAA 1035
Db 396 CTGTACCAAGTGCAGTCACTGTCGCGAGCTCGGCTCAATGCTCTTGTGACAGAGAA 455
Db 1036 GGGAGTGGGGGCTGTGTGCTCTCGCGCTCCAAAGCCAGGGTACCGGACTCACTGC 1095
Db 456 GGGAGTGGGGGCTGTGTGCTCTCGCGCTCCAAAGCCAGGGTACCGGACTCACTGC 515
Db 1096 GAGGAGATGGGCTTCTCAGGGGACAGACCACTCCGAGCTGAGCGTGGCGAGCGGGGC 1155
Db 516 GAGGAGATGGGCTTCTCAGGGGACAGACCACTCCGAGCTGAGCGTGGCGAGCGGGGC 575
Db 1156 GCCAATGCAAGTGGGCTTCTTCTGTGTGAGACAGAGGAGAGCTGCCACACCCAGAGG 1215
Db 576 GCCAATGCAAGTGGGCTTCTTCTGTGTGAGACAGAGGAGAGCTGCCACACCCAGAGG 635

DB 696 CAAAGCTGTGGCCGAGAGAGCTGCCCGTGGACCCGCTGTGGAGAGCCGGAGACACACG 755
QY 1336 TTGGGCGGTGGCCCTGGCAATCAGCTTGGCTATGATGAGACACACCTCTGTGGGGA 1395
DB 756 TTGGGCGGTGGCCCTGGCAATCAGCTTGGCTATGATGAGACACACCTCTGTGGGGA 815
QY 1396 TCCCTGTCTCCGGGAGCTGGGTGTGACAGCCGCGCCACTGTCTTCCGGAGGGAACGG 1455
DB 816 TCCCTGTCTCCGGGAGCTGGGTGTGACAGCCGCGCCACTGTCTTCCGGAGGGAACGG 875
QY 1456 GTCTGTCTCCGGGAGCTGGGTGTGACAGCCGCGCCACTGTCTTCCGGAGGGAACGG 1515
DB 876 GTCTGTCTCCGGGAGCTGGGTGTGACAGCCGCGCCACTGTCTTCCGGAGGGAACGG 935
QY 1516 CAGCTGGGGGTGACAGCTGTGGTCTACACCGGGGCTATCTTCCCTTCCGGAGGGAAC 1575
DB 936 CAGCTGGGGGTGACAGCTGTGGTCTACACCGGGGCTATCTTCCCTTCCGGAGGGAAC 995
QY 1576 AGCGAGGAGAACAGACGATTTGCTGTGCTACCTCTCCAGTCCCTTCCCTTCCAG 1635
DB 996 AGCGAGGAGAACAGACGATTTGCTGTGCTACCTCTCCAGTCCCTTCCCTTCCAG 1055
QY 1636 GAATATCATCCAGCTGTGTGCTCCAGCTGCGGCGCAGGCGCTGTGATGAGCAAGATC 1695
DB 1056 GAATATCATCCAGCTGTGTGCTCCAGCTGCGGCGCAGGCGCTGTGATGAGCAAGATC 1115
QY 1696 TGTATCGTACCGGGGTGGGGCAACGAGTACTATGAGCAACAGCGGGGTACTCCAG 1755
DB 1116 TGTATCGTACCGGGGTGGGGCAACGAGTACTATGAGCAACAGCGGGGTACTCCAG 1175
QY 1756 GAGCTCGAGTCCCATATACGCAATGATGTCTGCAATGAGGCTGACTTCTATGGAAC 1815
DB 1176 GAGCTCGAGTCCCATATACGCAATGATGTCTGCAATGAGGCTGACTTCTATGGAAC 1235
QY 1816 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGGGTGATGATGAGTCCGAG 1875
DB 1236 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGGGTGATGATGAGTCCGAG 1295
QY 1876 GAGCGACAGCGGTGTCTCTTGTGTGTGAGGACAGCATCTCTCGAGCCGACGTTGGGG 1935
DB 1296 GAGCGACAGCGGTGTCTCTTGTGTGTGAGGACAGCATCTCTCGAGCCGACGTTGGGG 1355
QY 1936 CTGTGTGCTGATGTGATGTTGGGGCACTGCTGTGTGCTGTGCTGTGCTGTGCTGTAC 1415
DB 1356 CTGTGTGCTGATGTGATGTTGGGGCACTGCTGTGTGCTGTGCTGTGCTGTGCTGTAC 1415
QY 1996 ACCAAAGTCAGTACTTCCGGGAGTGTACTTCCAGGCGATTAAGTCACTCCGAGAGCC 2055
DB 1416 ACCAAAGTCAGTACTTCCGGGAGTGTACTTCCAGGCGATTAAGTCACTCCGAGAGCC 1475
QY 2056 AGCGGATGTGACCCAGCTGTGACCGGGTGTCTTCTGCTGCGACCTCCAGAGGCGCGA 2115
DB 1476 AGCGGATGTGACCCAGCTGTGACCGGGTGTCTTCTGCTGCGACCTCCAGAGGCGCGA 1535
QY 2116 GGTATATCCCGGTGTGGGATCCACGCTGCGCGAGAGTGGAGCGTTTTTCTTCTGGGCC 1595
DB 1536 GGTATATCCCGGTGTGGGATCCACGCTGCGCGAGAGTGGAGCGTTTTTCTTCTGGGCC 1595
QY 2176 CCGTTCACAGGTCGCAAGGACACCTCCCTCCAGGGGTCTCTTCCAGTGGGGGGGCC 2235
DB 1596 CCGTTCACAGGTCGCAAGGACACCTCCCTCCAGGGGTCTCTTCCAGTGGGGGGGCC 1655
QY 2236 ACTCAGCCCGAGACACCAACCTCACCTCTCTGACCCCGCATGTAATATTTGTTCTGCT 2295
DB 1656 ACTCAGCCCGAGACACCAACCTCACCTCTCTGACCCCGCATGTAATATTTGTTCTGCT 1715
QY 2296 GTCTGGGACTCTGTCTAGAGTCCCTGTGATGATGAGTCTTTAAATTAATTAAGATGG 2355
DB 1716 GTCTGGGACTCTGTCTAGAGTCCCTGTGATGATGAGTCTTTAAATTAATTAAGATGG 1775
QY 2356 TTTTGATT 2363

DB 1776 TTTTGATT 1783
RESULT 13
US-10-295-027-1160
; Sequence 1160, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afari, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hezezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 60/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1160
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-295-027-1160
Query Match 67.4%; Score 1593.6; DB 17; Length 1783;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
QY 736 CCCCCAGCCCTGCTCCCGTCCATCTCTCAGAGGTCCACCTTGGCCAGAGGTCA 795
DB 157 CCTGAGGCTCCGCCCCACCTGCTGAGACCCA-GGGTCCACCTCGGCCAGAGGTCA 215
QY 796 GCCAGGATATTAACAAGAGGAGTGAATGGCGGAGAGAGGGTGGCGGAGCTGTG 855
DB 216 GCCAGGATATTAACAAGAGGAGTGAATGGCGGAGAGAGGGTGGCGGAGCTGTG 275
QY 856 CCATGCTGTCCAGAGCCAGAGTGGAGCTTCACTGCGGGACCTTGTACTTCTGACA 915
DB 276 CCATGCTGTCCAGAGCCAGAGTGGAGCTTCACTGCGGGACCTTGTACTTCTGACA 335
QY 916 GCCATGCGGGGCGGATCTTGGGCGATTTGTGCTTCTCTGAGAGTGAACAGAGCCG 975
DB 336 GCCATGCGGGGCGGATCTTGGGCGATTTGTGCTTCTCTGAGAGTGAACAGAGCCG 395
QY 976 CTGTATCCAGTGAAGTCAAGTCTGTGCGGAGCGCTCGCTCATGTGTTTGAACAACGGA 1035

Db CTGATCCAGTGTGACAGTCTGAGCTGCGACGCTCGCTCATGTCTTTTGACAAAGCGGAA 455
Oy 1036 GGAAGTGTGGCGGCTGTGTCTCTCGGCTCCAAAGCGGAGTATGCGGACTCAGCTGC 1095
Db 456 GGGAGTGTGGCGGCTGTGTCTCTCGGCTCCAAAGCGGAGTATGCGGACTCAGCTGC 515
Oy 1096 GAGGAGTGTGGCTTCTCAGGAGCACTGACCTCCGAGCTGAGACGTGTGAAACGGCGGGC 1155
Db 516 GAGGAGTGTGGCTTCTCAGGAGCACTGACCTCCGAGCTGAGACGTGTGAAACGGCGGGC 575
Oy 1156 GCCAATGACAGTGTGGGCTTCTGTGTGTGACGAGGAGGAGCTGCGCCACACCGAGAG 1215
Db 576 GCCAATGACAGTGTGGGCTTCTGTGTGTGACGAGGAGGAGCTGCGCCACACCGAGAG 635
Oy 1216 CTGCTGAGAGTCTATCTCGTGTGTGATTTGCCAGAGGCGTTTCTTGGCGGCACTGCG 1275
Db 636 CTGCTGAGAGTCTATCTCGTGTGTGATTTGCCAGAGGCGTTTCTTGGCGGCACTGCG 695
Oy 1276 CAAGCTGTGGCCGAGGAGCTGCGCGTGGACGCGATCGTGGAGGCGGAGGACACCGAGC 1335
Db 696 CAAGCTGTGGCCGAGGAGCTGCGCGTGGACGCGATCGTGGAGGCGGAGGACACCGAGC 755
Oy 1336 TTGGGCGCGGTGCGCTGTGCAAGTCAAGCTTGTGTATGATGAGACACCTCTGTGTGGGGA 1395
Db 756 TTGGGCGCGGTGCGCTGTGCAAGTCAAGCTTGTGTATGATGAGACACCTCTGTGTGGGGA 815
Oy 1396 TCCCTGCTCTCCGGGGAATGGGTGTGTGACAGCCGCCACTGCTTCCGGAGCGGAAACCG 1455
Db 816 TCCCTGCTCTCCGGGGAATGGGTGTGTGACAGCCGCCACTGCTTCCGGAGCGGAAACCG 875
Oy 1456 GTCCGTGTCCGATGTGGCGAGTGTGTTGGCGGTGCGGTGCGGCGGCGGCGGCGGCGGCGG 1515
Db 876 GTCCGTGTCCGATGTGGCGAGTGTGTTGGCGGTGCGGTGCGGCGGCGGCGGCGGCGGCGG 935
Oy 1516 CAGCTGTGGGAGTGCAGGCTGTGTGTACACAGGAGGCTATCTTCCCTTGTGGGAGCCCAAC 1575
Db 936 CAGCTGTGGGAGTGCAGGCTGTGTGTACACAGGAGGCTATCTTCCCTTGTGGGAGCCCAAC 995
Oy 1576 AGCGAGGAGAAACAGCAACGATATGCTGCTGCTCCAGCTGTGCTGCTGCTGCTGCTGCA 1635
Db 996 AGCGAGGAGAAACAGCAACGATATGCTGCTGCTCCAGCTGTGCTGCTGCTGCTGCTGCA 1055
Oy 1636 GAATACATCCAGGCTGTGTGCTCCAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1695
Db 1056 GAATACATCCAGGCTGTGTGCTCCAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1115
Oy 1696 TGTACCGTGAAGGCTGTGGGCAACGCGAGTACTATGTGCAACAGGCGGAGGTACTCCAG 1755
Db 1116 TGTACCGTGAAGGCTGTGGGCAACGCGAGTACTATGTGCAACAGGCGGAGGTACTCCAG 1175
Oy 1756 GAGGCTCGAGTCCCATATATGCAATATGTGTGCAATGTGCGCTGACTTCTTATGGAAC 1815
Db 1176 GAGGCTCGAGTCCCATATATGCAATATGTGTGCAATGTGCGCTGACTTCTTATGGAAC 1235
Oy 1816 CAGATCAAGCCCAAGATGTCTGTGCTGTGCTGTGAGGACAGCATCTGTGACGCGCACTTGGCGG 1875
Db 1236 CAGATCAAGCCCAAGATGTCTGTGCTGTGCTGTGAGGACAGCATCTGTGACGCGCACTTGGCGG 1295
Oy 1876 GGGCAACAGCGGTGTCTCTTGTGTGTGTGAGGACAGCATCTGTGACGCGCACTTGGCGG 1935
Db 1296 GGGCAACAGCGGTGTCTCTTGTGTGTGTGAGGACAGCATCTGTGACGCGCACTTGGCGG 1355
Oy 1936 CTGTGTGTGATTTGT 1995
Db 1356 CTGTGTGTGATTTGT 1415
Oy 1996 ACCAAAGTCAAGTACTTCCGGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2055
Db 1416 ACCAAAGTCAAGTACTTCCGGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1475
Oy 2056 AGCGGATGTGATCCAGCTGTGACGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2115

Db 1476 AGCGGATGTGATCCAGCTCTGTGACGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1535
Oy 2116 GTGTATCCCGGT 2175
Db 1536 GTGTATCCCGGT 1595
Oy 2176 CGGTCCAGAGTCCCAAGGACACCTCTCCAGGAGTCTCTCTTCCAGTGTGCGGCGCC 2235
Db 1596 CGGTCCAGAGTCCCAAGGACACCTCTCCAGGAGTCTCTCTTCCAGTGTGCGGCGCC 1655
Oy 2236 ACTCAGCCCGGAGACCAACCACTCACTCTGTGACCTCCCAAGTAAATATTTGTGTGT 2295
Db 1656 ACTCAGCCCGGAGACCAACCACTCACTCTGTGACCTCCCAAGTAAATATTTGTGTGT 1715
Oy 2296 GTCTGTGAGATCTGT 2355
Db 1716 GTCTGTGAGATCTGT 1775
Oy 2356 TTTTGATT 2363
Db 1776 TTTTGATT 1783

RESULT 14
US-10-156-214A-34
; Sequence 34, Application US/10156214A
; Publication No. US20040001801A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Joseph Edward Sample
; APPLICANT: George P. Vlasuk
; APPLICANT: Scott Jeffrey Kemp
; APPLICANT: Mallareddy Komandla
; APPLICANT: Daniel Varma Slev
; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic U
; FILE REFERENCE: 24745-1611
; CURRENT FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Homo Sapien
; NAME/KEY: CDS
; LOCATION: (246)...(1499)
; OTHER INFORMATION: Nucleic acid encoding human hepsin
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank M18930
; DATABASE ENTRY DATE: 1993-06-11
US-10-156-214A-34

Query Match 67.4%; Score 1593.6; DB 17; Length 1783;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

Oy 736 CCCCCAGGCCCTGCTCCCGTCCATCTCTCAAGGTCCCACTGAGCCAGAGAGTCA 795
Db 157 CTTGAGAGCTCCGCCCCCACTGTGTGAGCCCA-GGGTCCACCTGTGCGCCAGAGAGTCA 215
Oy 796 GCCAGGGAATCATTAACAAGGAGGAGCATGAGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAG 855
Db 216 GCCAGGGAATCATTAACAAGGAGGAGGAGCATGAGCGGAGAGAGGAGGAGGAGGAGGAGGAG 275
Oy 856 CCAATGTCTCTCAGACCCCAAGTGTGAGCTCTCACTGCGGAGACCTGTCTACTTCTTGACA 915
Db 276 CCAATGTCTCTCAGACCCCAAGTGTGAGCTCTCACTGCGGAGACCTGTCTACTTCTTGACA 335
Oy 916 GCCATGTGGGCGGAGTCTGTGGGCGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 975
Db 336 GCCATGTGGGCGGAGTCTGTGGGCGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 395

Db 456 GGGAGTGGCGGCTGCTGTGCTCTCCGCGCTCCAAACGCAAGGTAGCCGAGCTCAGCTGC 515
Qy 1096 GAGAGATGGGCTTCTCAGGCACTGACCACTCCGAGCTGAGCTGCAACGGCGGAC 1155
Db 516 GAGAGATGGGCTTCTCAGGCACTGACCACTCCGAGCTGAGCTGCAACGGCGGAC 575
Qy 1156 GCCATGAGCAGTGGGCTTCTTGTGTGAGACGAGGAGGCTGCCCCACACCCAGAG 1215
Db 576 GCCATGAGCAGTGGGCTTCTTGTGTGAGACGAGGAGGCTGCCCCACACCCAGAG 635
Qy 1216 CTGCTGAGGATCATCTCCGTGTGTGATTTGCCAGAGGCGCTTCTTGGCCGCACTGCG 1275
Db 636 CTGCTGAGGATCATCTCCGTGTGTGATTTGCCAGAGGCGCTTCTTGGCCGCACTGCG 695
Qy 1276 CAAGACTGTGGCCGAGGAGCTGCCCTGAGACCGCATGTGGAGGCGGAGACACAGC 1335
Db 696 CAAGACTGTGGCCGAGGAGCTGCCCTGAGACCGCATGTGGAGGCGGAGACACAGC 755
Qy 1336 TTGGGCGGATGGCGGTGGCAAGTCAAGCTTGCTATGATGAGACACACTGTGTGGGGA 1395
Db 756 TTGGGCGGATGGCGGTGGCAAGTCAAGCTTGCTATGATGAGACACACTGTGTGGGGA 815
Qy 1396 TCCCTGCTCTCCGAGGACTGGGTGCTGACAGCCGCGCACTGCTTCCCGAGCGGAAACCG 1455
Db 816 TCCCTGCTCTCCGAGGACTGGGTGCTGACAGCCGCGCACTGCTTCCCGAGCGGAAACCG 875
Qy 1456 GTCTGTCTCCGATGGCGAGTGTGGCCGCTGCGGTCCGAGGCTTCCCGACGGTCTG 1515
Db 876 GTCTGTCTCCGATGGCGAGTGTGGCCGCTGCGGTCCGAGGCTTCCCGACGGTCTG 935
Qy 1516 CAGCTGGGGGGTGCAGGCTGTGTGTATCCAGCGGGGCTATCTTCCCTTTCGAGGACCCCAAC 1575
Db 936 CAGCTGGGGGGTGCAGGCTGTGTGTATCCAGCGGGGCTATCTTCCCTTTCGAGGACCCCAAC 995
Qy 1576 AGCGAGGAGAACAGCAACGATATTGCCCTGTGTCACCTCTCCAGTCCCTGCCCCCTCACA 1635
Db 996 AGCGAGGAGAACAGCAACGATATTGCCCTGTGTCACCTCTCCAGTCCCTGCCCCCTCACA 1055
Qy 1636 GAATATCATCCAGCTGTGTGCTCCAGCTGCCGAGCGCTGTGTGATGAGCAAGATC 1695
Db 1056 GAATATCATCCAGCTGTGTGCTCCAGCTGCCGAGCGCTGTGTGATGAGCAAGATC 1115
Qy 1696 TGTACCGTGAAGGGCTGGGGCAACAGCAGTACTATGGCAACAGGCGGGGTACTCCAG 1755
Db 1116 TGTACCGTGAAGGGCTGGGGCAACAGCAGTACTATGGCAACAGGCGGGGTACTCCAG 1175
Qy 1756 GAGGCTCGAGTCCCATATATCAGCATATGATGTGCAATGAGCGCTGACTTCTATGAAAC 1815
Db 1176 GAGGCTCGAGTCCCATATATCAGCATATGATGTGCAATGAGCGCTGACTTCTATGAAAC 1235
Qy 1816 CAGATCAAGGCCAAGATGTCTGTGTGTGCTAACCCCGAGGGTGGCATTTGATGCTGCCAG 1875
Db 1236 CAGATCAAGGCCAAGATGTCTGTGTGTGCTAACCCCGAGGGTGGCATTTGATGCTGCCAG 1295
Qy 1876 GGCGACAGCGGTGGTCCCTTGTGTGTGAGACAGCATCTCCGAGCGCCAGTTGGCGG 1935
Db 1296 GGCGACAGCGGTGGTCCCTTGTGTGTGAGACAGCATCTCCGAGCGCCAGTTGGCGG 1355
Qy 1936 CTGTGTGAGCATTTGATTTGGGGCACTGGCTGTGCCCTGGCCAGAAACAGGCGCTTAC 1995
Db 1356 CTGTGTGAGCATTTGATTTGGGGCACTGGCTGTGCCCTGGCCAGAAACAGGCGCTTAC 1415
Qy 1996 ACCAAAGTCAAGTCTTCCGAGAGTGGATCTTCCAGGCGCATTAAGACTCACTCCGAGGCC 2055
Db 1416 ACCAAAGTCAAGTCTTCCGAGAGTGGATCTTCCAGGCGCATTAAGACTCACTCCGAGGCC 1475
Qy 2056 AGCGGCAATGTGATCCAGCTGACCGGTGCTTCTGCTGCGGAGCTCCAGGGCCCGA 2115
Db 1476 AGCGGCAATGTGATCCAGCTGACCGGTGCTTCTGCTGCGGAGCTCCAGGGCCCGA 1535
Qy 2116 GGTGATCCCGGTGTGGATCCAGCTTGGCCGAGGATGGAGCGTTTCTTCTTGGGCGC 2175

Db 1536 GGTGATCCCGGTGTGGATCCAGCTGGGCCGAGATGGGACGTTTTTCTTGTGGGCGC 1595
Qy 2176 CGGTCCACAGTCCAGAGACACCTCCCTCAGGCTCCTCTTCCACAGTGGGGGCGCC 2235
Db 1596 CGGTCCACAGTCCAGAGACACCTCCCTCAGGCTCCTCTTCCACAGTGGGGGCGCC 1655
Qy 2236 ACTGAGCCCGAGACACCCCAACTCACTCTGAGCCCCCATGTAAATATTTGTGCT 2295
Db 1656 ACTGAGCCCGAGACACCCCAACTCACTCTGAGCCCCCATGTAAATATTTGTGCT 1715
Qy 2296 GTCTGGAGCTCTGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2355
Db 1716 GTCTGGAGCTCTGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1775
Qy 2356 TTTTGATT 2363
Db 1776 TTTTGATT 1783

Search completed: August 17, 2005, 17:58:19
Job time : 4019 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2005, 01:25:31 ; Search time 16703 Seconds
(without alignments)
6855.037 Million cell updates/sec

Title: US-10-054-498-1
Perfect score: 2363
Sequence: 1 tcgagccgccttcacagga.....taataaagatgtttgac 2363

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_hg:*
2: gb_bra:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sbs:*
12: gb_sy:*
13: gb_un:*
14: gb_vt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2363	100.0	2363	6	AR220138
2	2363	100.0	2363	6	AX336397 Sequence
3	2363	100.0	2363	6	AX411046
4	2363	100.0	2363	6	AX511632 Sequence
5	2363	100.0	2363	9	HSHEPSH
6	1593.6	67.4	1783	6	CO718142 Sequence
7	1593.6	67.4	1783	6	AR280046
8	1593.6	67.4	1783	6	AX207963
9	1593.6	67.4	1783	9	HUMHPSNA
10	1592	67.4	1761	9	BC025716
11	1405	59.5	1769	6	CO771486
12	1269	53.7	1615	6	AR255885
13	1199	50.7	1199	9	HSHEPSL
14	1023.6	43.3	1739	10	RNHRPA
15	1022.2	43.3	1781	10	AF030065
16	1008.2	42.7	1605	6	AK085656
17	906.2	38.3	1887	10	AY234104
18	858.4	36.3	2175	6	AX833435
19	858.4	36.3	2175	9	AK095160

20	806.8	34.1	169891	9	AC020907
21	779.6	33.0	1580	10	BC072688
22	600	25.4	600	6	AX530376
23	391.2	16.6	614	6	AR263893
24	380.2	16.1	497	6	AX698852
25	325	13.8	21784	6	AR255886
26	325	13.6	174875	2	AC024682
27	321.8	13.6	2547	6	AX747294
28	321.8	13.6	2547	9	AK091988
29	297	12.6	297	11	G06603
30	249.8	10.6	266	6	AX335256
31	249.8	10.6	266	6	AX409165
32	235	9.9	2831	6	CO842731
33	235	9.9	2831	9	AK125670
34	192	8.1	601	6	AR255897
35	191	8.1	601	6	AR255898
36	170	7.2	133276	2	AC087147
37	170	7.2	216120	2	AC021477
38	169.4	7.2	601	6	AR255896
39	168.4	7.1	239049	2	AC118870
40	168.4	7.1	244224	2	AC125992
41	163.8	6.9	177	1	GI4606
42	150.8	6.4	2237	9	AY190317
43	148.4	6.3	1078	6	AX190545
44	148.4	6.3	1314	6	AX190543
45	148.4	6.3	1314	6	AX190546

ALIGNMENTS

RESULT 1	AR220138	Sequence 3 from patent US-6423543.	DNA	Linear	PAT 26-SEP-2002
LOCUS	AR220138				
DEFINITION	Sequence 3 from patent US-6423543.				
ACCESSION	AR220138				
VERSION	AR220138.1	GI:23574581			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 2363)				
AUTHORS	Marcotte, P.A. and Cowser, L.M.				
TITLE	Antisense modulation of heparin expression				
JOURNAL	Patent: US-6423543-A 3 23-JUL-2002;				
FEATURES	Location/Qualifiers				
source	1. 2363				
	/mol_type="genomic DNA"				
ORIGIN					
Query Match	Best Local Similarity 100.0%; Score 2363; DB 6; Length 2363;				
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	TCGAGCCCGCTTTCAGGAGACCTTACCTGAGGGCCACAGGTAGAGCAGCTGAGCTAGC	60		
DB	1	TCGAGCCCGCTTTCAGGAGACCTTACCTGAGGGCCACAGGTAGAGCAGCTGAGCTAGC	60		
QY	61	AGGCCCCAGCCACCGCTGCTGCTCAGGCGCCCGCTGCTGCGGGCCACATGCTCC	120		
DB	61	AGGCCCCAGCCACCGCTGCTGCTCAGGCGCCCGCTGCTGCGGGCCACATGCTCC	120		
QY	121	TGCGCAGGCTTGAAGTACGAGCCGAGCCTTACCTGAGGCTCCGCCACCTAGC	180		
DB	121	TGCGCAGGCTTGAAGTACGAGCCGAGCCTTACCTGAGGCTCCGCCACCTAGC	180		
QY	181	TGAGCCCGAGGTTAAGAGCAGGGCCCGAGCTCAGAGTTCCAGCCCTTGAAGACAGGG	240		
DB	181	TGAGCCCGAGGTTAAGAGCAGGGCCCGAGCTCAGAGTTCCAGCCCTTGAAGACAGGG	240		
QY	241	TTCCCTCATCCCCCAGCCCGAGCTTAATGACCTCTTAATGAGGGGTTCTCGGGAGC	300		

Db	241	TTTCCCTAATCCCCCAACCCAGCCTAATGCCACCTCTCTAATATAGAGGGGTTCTGGGGAAC	300
QY	301	TGAAGAGGGGGCATTATGACGTCTCCCAAGCACCAGGTGTTCTGCTGCTCTT	360
Db	301	TGAAGAGGGGGCATTATGACGTCTCCCAAGCACCAGGTGTTCTGCTGCTCTT	360
QY	361	CAGACTCAGCCGTTGGACCCCAAGTCTTTCTCTCCCAAGCACCAGAGTTCCAGCCCTCAG	420
Db	361	CAGACTCAGCCGTTGGACCCCAAGTCTTTCTCTCCCAAGCACCAGAGTTCCAGCCCTCAG	420
QY	421	GGCCCTCCCTCAATPACTAGGGAAGTCGGGGCCCCAAATTCCTCTTTCCCAAGACTTA	480
Db	421	GGCCCTCCCTCAATPACTAGGGAAGTCGGGGCCCCAAATTCCTCTTTCCCAAGACTTA	480
QY	481	TGATTTCAAGTCTCAGCTGTCCTCTCTCTCAAACCGGGANTCTCAGTCCCTGCTCCAC	540
Db	481	TGATTTCAAGTCTCAGCTGTCCTCTCTCTCAAACCGGGANTCTCAGTCCCTGCTCCAC	540
QY	541	CAGGCTCAGGCATGGGGGGTCCCCATTCCTTGCAATTCAGGCGTCCCCCGCTGTGTCA	600
Db	541	CAGGCTCAGGCATGGGGGGTCCCCATTCCTTGCAATTCAGGCGTCCCCCGCTGTGTCA	600
QY	601	GACACTGACCCCATCTTGAAACCCAGCCCAATCTGTCCTGGATCAAGGCTGTCTGG	660
Db	601	GACACTGACCCCATCTTGAAACCCAGCCCAATCTGTCCTGGATCAAGGCTGTCTGG	660
QY	661	CCAAAGGCCAGTCCCTTACAGCTGTGCTGGATGGAAGCCTGGAGCTGGGGGCGCAGAGCT	720
Db	661	CCAAAGGCCAGTCCCTTACAGCTGTGCTGGATGGAAGCCTGGAGCTGGGGGCGCAGAGCT	720
QY	721	GGGCTGGGCTGGGCTTCCCCAGGCGCTGTCTCCCGTCAATCTCTCAAGGTCCACCC	780
Db	721	GGGCTGGGCTGGGCTTCCCCAGGCGCTGTCTCCCGTCAATCTCTCAAGGTCCACCC	780
QY	781	TGGCCCAAGAGGTGACGAGGAATCATTTAACAAGCAGTGAATGGCCGAGAAAGAG	840
Db	781	TGGCCCAAGAGGTGACGAGGAATCATTTAACAAGCAGTGAATGGCCGAGAAAGAG	840
QY	841	GATGGCCGGACGTGTCACATGTGCTTCCAGACCAGAGTGGCAGCTCTCACTGCGGGGACC	900
Db	841	GATGGCCGGACGTGTCACATGTGCTTCCAGACCAGAGTGGCAGCTCTCACTGCGGGGACC	900
QY	901	CTGCTACTTTCGACAGCCATGGGGCGGCGATCTTGGGCCAATTGTGGCTGTCTCTCAGG	960
Db	901	CTGCTACTTTCGACAGCCATGGGGCGGCGATCTTGGGCCAATTGTGGCTGTCTCTCAGG	960
QY	961	AGTAGCCAGAGGCCGCTGTACCCCAATGACAGTCACTCTGCGGAGAGCTCGCTCATGTGTC	1020
Db	961	AGTAGCCAGAGGCCGCTGTACCCCAATGACAGTCACTCTGCGGAGAGCTCGCTCATGTGTC	1020
QY	1021	TTTTCACAAGACGGAAGGACGTGGGGGCTGTGTCCTCTCGCGCTCCAAAGCCAGGGTA	1080
Db	1021	TTTTCACAAGACGGAAGGACGTGGGGGCTGTGTCCTCTCGCGCTCCAAAGCCAGGGTA	1080
QY	1081	GCCGAGCTCACTGCGAGAGATGGGCTTTCCTCAGGGGACCTGACCACCTCCGAGCTGAGC	1140
Db	1081	GCCGAGCTCACTGCGAGAGATGGGCTTTCCTCAGGGGACCTGACCACCTCCGAGCTGAGC	1140
QY	1141	GTGCGCAATGGCGGGGGCCATGGCACGTGCGGCTTCTTCTGTGTGGAAGAGGGGAGGCTG	1200
Db	1141	GTGCGCAATGGCGGGGGCCATGGCACGTGCGGCTTCTTCTGTGTGGAAGAGGGGAGGCTG	1200
QY	1201	CCCCACACCCAGAGGCTGTGGAGGTCAATCTCCGTGTGATTTGGCCCAAGAGCGGTTTC	1260
Db	1201	CCCCACACCCAGAGGCTGTGGAGGTCAATCTCCGTGTGATTTGGCCCAAGAGCGGTTTC	1260
QY	1261	TTGGCGGCAATTCGCAAGACTGTGGCCGCAAGAGACTGCCGTGGAACCGCATCTGTGGGA	1320
Db	1261	TTGGCGGCAATTCGCAAGACTGTGGCCGCAAGAGACTGCCGTGGAACCGCATCTGTGGGA	1320
QY	1321	GGCGGGGACACCAAGCTTGGGCGGTGGCCGTGGCAAGTCAAGCTTCTCGTATGATGGAGCA	1380
Db	1321	GGCGGGGACACCAAGCTTGGGCGGTGGCCGTGGCAAGTCAAGCTTCTCGTATGATGGAGCA	1380

QY	1381	CACCTCTGTGGGGGATCCCTGCTCTTCGGGGAGCTGGGTGTGACAGCCGCTCACTGCTTC	1440
Db	1381	CACCTCTGTGGGGGATCCCTGCTCTTCGGGGAGCTGGGTGTGACAGCCGCTCACTGCTTC	1440
QY	1441	CCGGAGCCGGAACCCGGGTCCGTGCCGATGGCCAGTGTTCGGGTGCCGGTCCGAGGCC	1500
Db	1441	CCGGAGCCGGAACCCGGGTCCGTGCCGATGGCCAGTGTTCGGGTGCCGGTCCGAGGCC	1500
QY	1501	TCTCCCAACGATCTGACGTGGGGGTGACAGCTGTGTCTACCAACGGGGCTATCTTCCC	1560
Db	1501	TCTCCCAACGATCTGACGTGGGGGTGACAGCTGTGTCTACCAACGGGGCTATCTTCCC	1560
QY	1561	TTTGGGGACCCCAACAGCGAGGAGAAACAGCAACGATATTGCCCTGTGCTCACTCTCCAGT	1620
Db	1561	TTTGGGGACCCCAACAGCGAGGAGAAACAGCAACGATATTGCCCTGTGCTCACTCTCCAGT	1620
QY	1621	CCCCCTCCCTCAACGAATACATCCAGCCCTGTGTCTCCCACTGCGGGCCAGGCCCTG	1680
Db	1621	CCCCCTCCCTCAACGAATACATCCAGCCCTGTGTCTCCCACTGCGGGCCAGGCCCTG	1680
QY	1681	GTGATGTGGCAAGATCTGTACCGGTACGGGCTGGGGCAACGCAAGTACTATGTGCCAAAG	1740
Db	1681	GTGATGTGGCAAGATCTGTACCGGTACGGGCTGGGGCAACGCAAGTACTATGTGCCAAAG	1740
QY	1741	GCCGGGGGTACTCAGAGGGCTGAGTCCCATTAATCAGCAATGATGTTCGAATGCGCT	1800
Db	1741	GCCGGGGGTACTCAGAGGGCTGAGTCCCATTAATCAGCAATGATGTTCGAATGCGCT	1800
QY	1801	GACTTCTATGAAACCAATCAAGCCCAAGATGTTCGTGTCTGTGCTAACCCCGAGGGTGGC	1860
Db	1801	GACTTCTATGAAACCAATCAAGCCCAAGATGTTCGTGTCTGTGCTAACCCCGAGGGTGGC	1860
QY	1861	ATTGATGCTGTGCAGGGCGACAGCGGTGTCCCTTGTGTGTGAGGACAGCATCTCGG	1920
Db	1861	ATTGATGCTGTGCAGGGCGACAGCGGTGTCCCTTGTGTGTGAGGACAGCATCTCGG	1920
QY	1921	ACGCCACGTTGGCGGCTGTGTGACATTTGTAGTTGGGGCACTGGCTGTGCCCTGGCCAG	1980
Db	1921	ACGCCACGTTGGCGGCTGTGTGACATTTGTAGTTGGGGCACTGGCTGTGCCCTGGCCAG	1980
QY	1981	AAGCCAGGCGTATAACCAAAATCACTGATCTTCGGGAGATGATCTTCCAGGCGCATPAAG	2040
Db	1981	AAGCCAGGCGTATAACCAAAATCACTGATCTTCGGGAGATGATCTTCCAGGCGCATPAAG	2040
QY	2041	ACTCACTCCGAAGCCAGCGGATGGGATCCAGACTGTGACCGGTGTCTTCGCTCGCA	2100
Db	2041	ACTCACTCCGAAGCCAGCGGATGGGATCCAGACTGTGACCGGTGTCTTCGCTCGCA	2100
QY	2101	GCCTCCAGGCGCCGAGGTGATCCGGTGTGTGGATTCACGCTGGGCGAGAGATGGAGCT	2160
Db	2101	GCCTCCAGGCGCCGAGGTGATCCGGTGTGTGGATTCACGCTGGGCGAGAGATGGAGCT	2160
QY	2161	TTTTTCTTCTTGGGCCCGGTTCACAGGTCCAAAGACACCTCTCTCCAGGGTCTCTCTTC	2220
Db	2161	TTTTTCTTCTTGGGCCCGGTTCACAGGTCCAAAGACACCTCTCTCTCCAGGGTCTCTCTTC	2220
QY	2221	CACAGTGGGGGCGCACTCAGCCCGAGACCAACCACTCACTCTGACCCGCAATGT	2280
Db	2221	CACAGTGGGGGCGCACTCAGCCCGAGACCAACCACTCACTCTGACCCGCAATGT	2280
QY	2281	AAATATTGTTCTGCTGTCTGGACTCTGTCTTAGGTGCCCTGATGATGGAGTCTCTTT	2340
Db	2281	AAATATTGTTCTGCTGTCTGGACTCTGTCTTAGGTGCCCTGATGATGGAGTCTCTTT	2340
QY	2341	AAATATAAAGATGTTTGATT 2363	
Db	2341	AAATATAAAGATGTTTGATT 2363	

DEFINITION Sequence 6906 from Patent WO0194629.
 ACCESSION AX336397
 VERSION AX336397.1 GI:18127116
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Young, P.E., Augustus, M., Carter, K.C., Edner, R., Endress, C.,
 Horrigan, S., Soppel, D.R. and Weaver, Z.
 TITLE Cancer gene determination and therapeutic screening using signature
 gene sets
 JOURNAL Patent: WO 0194629-A 6906 13-DEC-2001;
 Avalon Pharmaceuticals (US)
 FEATURES
 Source 1. 2363
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Query Match 100.0%; Score 2363; DB 6; Length 2363;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

721 GGGCTGGGCTGGGGCTCCCCCAGAGCCCTGCTCCCCGTCATCTCTCAAGTCCACCC 780
 721 GGGCTGGGCTGGGGCTCCCCCAGAGCCCTGCTCCCCGTCATCTCTCAAGTCCACCC 780
 781 TGGCCCAAGAGGTGAGCCAGGGAATCTAATCAAGAGGAGTGAATGAGCCAGAGAG 840
 781 TGGCCCAAGAGGTGAGCCAGGGAATCTAATCAAGAGGAGTGAATGAGCCAGAGAG 840
 841 GGTGGCCGGAATGTGCAATGTGCTGCTCAAGACCAAGGTGAGAGTCTCACTGCGGAG 900
 841 GGTGGCCGGAATGTGCAATGTGCTGCTCAAGACCAAGGTGAGAGTCTCACTGCGGAG 900
 901 CTGCTACTTCTGAGAGGCAATCGGGGCGGCAATCTGCGGCAATTTGGCTGTTCTCTCAG 960
 901 CTGCTACTTCTGAGAGGCAATCGGGGCGGCAATCTGCGGCAATTTGGCTGTTCTCTCAG 960
 961 AGTACCAAGAGGCGGCTGTACCCAGTACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1020
 961 AGTACCAAGAGGCGGCTGTACCCAGTACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1020
 1021 TTGACCAAGAGGCGGCTGTACCCAGTACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1080
 1021 TTGACCAAGAGGCGGCTGTACCCAGTACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1080
 1081 GCGGAGCTCAAGTCTGAGAGGAGTGGGCTTCTTCAAGGCACTGACCTCCAGCTGAG 1140
 1081 GCGGAGCTCAAGTCTGAGAGGAGTGGGCTTCTTCAAGGCACTGACCTCCAGCTGAG 1140
 1141 GTGCAAGAGGCGGCGGCGCAATGGCAAGTGGGCGCTTCTTGTGTGAGAGAGGAGGCTG 1200
 1141 GTGCAAGAGGCGGCGGCGCAATGGCAAGTGGGCGCTTCTTGTGTGAGAGAGGAGGCTG 1200
 1201 CCCCACACCAAGAGGCTGTGAGAGTATCTCGTGTGTATTTGCCCCAGAGGCGCTTTC 1260
 1201 CCCCACACCAAGAGGCTGTGAGAGTATCTCGTGTGTATTTGCCCCAGAGGCGCTTTC 1260
 1261 TTGGCCGCAATCTGCAAGAGTGTGGCGGAGAGTGGCGGAGAGTGGCGGAGAGTGG 1320
 1261 TTGGCCGCAATCTGCAAGAGTGTGGCGGAGAGTGGCGGAGAGTGGCGGAGAGTGG 1320
 1321 GGGCGGAGCAAGAGTGTGGCGGAGAGTGGCGGAGAGTGGCGGAGAGTGGCGGAGAGT 1380
 1321 GGGCGGAGCAAGAGTGTGGCGGAGAGTGGCGGAGAGTGGCGGAGAGTGGCGGAGAGT 1380
 1381 CACTCTGTGGGAGATCTGCTCTCCGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1440
 1381 CACTCTGTGGGAGATCTGCTCTCCGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1440
 1441 CCGAGGCGGAACCGGAGTCTGTCTCCGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 1500
 1441 CCGAGGCGGAACCGGAGTCTGTCTCCGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 1500
 1501 TCTCCCAAGGAGTCTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1560
 1501 TCTCCCAAGGAGTCTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1560
 1561 TTTCCGAGAGGAGTCTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1620
 1561 TTTCCGAGAGGAGTCTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1620
 1621 CCGCTGCGCTTCAAGAGTCTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1680
 1621 CCGCTGCGCTTCAAGAGTCTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1680
 1681 GTGAGAGGAGAGTCTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1740
 1681 GTGAGAGGAGAGTCTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1740
 1741 GCGGAGGAGTCTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1800
 1741 GCGGAGGAGTCTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1800
 1741 GCGGAGGAGTCTGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1800

OY	1801	GACCTTCTATGGAAACCAAGATCAAGCCCAAGATGTTCTGTGCTGGCTACCCGAGGGTGGC	1860		
Db	1801	GACTTCTATGGAACCAAGATCAAGCCCAAGATGTTCTGTGCTGGCTACCCGAGGGTGGC	1860		
OY	1861	ATTGATGCTGTGCAGGGGCGACAGCGGGTGGTCCCTTGTGTGTGAAGAACAGATCTTCGCG	1920		
Db	1861	ATTGATGCTGTGCAGGGGCGACAGCGGGTGGTCCCTTGTGTGTGAAGAACAGATCTTCGCG	1920		
OY	1921	ACGCCACGTTGGCGGGCTGTGTGGCAATTGTGATTTGGGGCACTGGCTGTGCCCTGGCCAG	1980		
Db	1921	ACGCCACGTTGGCGGGCTGTGTGGCAATTGTGATTTGGGGCACTGGCTGTGCCCTGGCCAG	1980		
OY	1961	AAGCCAGGCGCTCTACCAACCAAGTCAGTACTTCGCGGAGTGGATCTTCAGGCCATTAAG	2040		
Db	1961	AAGCCAGGCGCTCTACCAACCAAGTCAGTACTTCGCGGAGTGGATCTTCAGGCCATTAAG	2040		
OY	2041	ACTCATCTCCGAAGCCAGGCGGATGGTGAACCAAGCTGTGACCGGGTGGCTTCTCGCTGGCA	2100		
Db	2041	ACTCATCTCCGAAGCCAGGCGGATGGTGAACCAAGCTGTGACCGGGTGGCTTCTCGCTGGCA	2100		
OY	2101	GCTTCAGAGGCGCCGAGTGTATCCCGGTGGTGGATTCACGCTGGGCGAGGATGGAGCT	2160		
Db	2101	GCTTCAGAGGCGCCGAGTGTATCCCGGTGGTGGATTCACGCTGGGCGAGGATGGAGCT	2160		
OY	2161	TTTTCTTCTTGGGCCCCGGTTCACAGGTCCAAAGACACCCCTCCTCAAGGTCCTCTCTTC	2220		
Db	2161	TTTTCTTCTTGGGCCCCGGTTCACAGGTCCAAAGACACCCCTCCTCAAGGTCCTCTCTTC	2220		
OY	2221	CACAGTGGCGGGCCCACTCAGGCCCGAGACCAACCCATCTCAACCTCTCTGACCCCATGT	2280		
Db	2221	CACAGTGGCGGGCCCACTCAGGCCCGAGACCAACCCATCTCAACCTCTCTGACCCCATGT	2280		
OY	2281	AAATATTTCTTCTGCTGCTGTGGGACCTCCCTGTCTAGTGTGCCCTGTGATGATGGATCTCTTT	2340		
Db	2281	AAATATTTCTTCTGCTGCTGTGGGACCTCCCTGTCTAGTGTGCCCTGTGATGATGGATCTCTTT	2340		
OY	2341	AAATATTAAGATGGTTTTGATT	2363		
Db	2341	AAATATTAAGATGGTTTTGATT	2363		
RESULT 3	AX411046	2363 bp	DNA	linear	PAT 14-JUN-2002
LOCUS	AX411046				
DEFINITION	Sequence 3693 from Patent WO0229103.				
ACCESSION	AX411046				
VERSION	AX411046.1	GI:21443751			
KEYWORDS	.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1				
AUTHORS	Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.				
TITLE	Gene expression profiles in liver cancer				
JOURNAL	Patent: WO 0229103-A 3693 11-APR-2002;				
	GENE LOGIC INC (US)				
FEATURES	Location/Qualifiers				
source	1..2363				
	/organism="Homo sapiens"				
	/mol_type="unassigned DNA"				
	/db_xref="caxon:9606"				
ORIGIN	/note="EMBL/GenBank Accession No. X07732"				

	Query Match	100.0%;	Score 2363;	DB 6;	Length 2363;	
	Best Local Similarity	100.0%;	Pred. No. 0;			
	Matches 2363;	Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;
Qy	1	TCGAGCCCCGCTTCCAGGGACCTTA	CCTGAGAGGCCCAAGGTGAGGCAGCCTGAGCCTAAGC	60		
Db	1	TCGAGCCCCGCTTCCAGGGACCTTA	CCTGAGAGGCCCAAGGTGAGGCAGCCTGAGCCTAAGC	60		

QY	61	AGGCCCCACGACCCGCTCTGCTCCAGAGCGCGCCGCTGCTGGGGGACCAATGCTCC	120
Db	61	AGGCCCCACGACCCGCTCTGCTCCAGAGCGCGCCGCTGCTGGGGGACCAATGCTCC	120
QY	121	TGCCAGGCTCGAGACTGACCCGACCCGGGCACTTACTCGAGGCTCCGCCCACTGTC	180
Db	121	TGCCAGGCTCGAGACTGACCCGACCCGGGCACTTACTCGAGGCTCCGCCCACTGTC	180
QY	181	TGACCCCAAGGTAAAGACAAGGGCCCCAGACTCAAGTTCCAGCCCTGAGACAAGGG	240
Db	181	TGACCCCAAGGTAAAGACAAGGGCCCCAGACTCAAGTTCCAGCCCTGAGACAAGGG	240
QY	241	TTCCCTCATCCCCCAACCCAGCTTAATGCCACTCTTAATAGAGGGGTTCTGGGACC	300
Db	241	TTCCCTCATCCCCCAACCCAGCTTAATGCCACTCTTAATAGAGGGGTTCTGGGACC	300
QY	301	TGAAGAGGGGCACTATGACGTCTCCCAAGACCTTAGTGTGTGCTCTTCTT	360
Db	301	TGAAGAGGGGCACTATGACGTCTCCCAAGACCTTAGTGTGTGCTCTTCTT	360
QY	361	CAGACTAGCCGTTGACCCCGAGTCTTTCTCCCGACAGCCAGAGTTCCAGCCTCAG	420
Db	361	CAGACTAGCCGTTGACCCCGAGTCTTTCTCCCGACAGCCAGAGTTCCAGCCTCAG	420
QY	421	GCCCCCTCTCCTCATCTAGAGGTCTTGCCCCCAATTCCTCTTTCCCAAGACTTA	480
Db	421	GCCCCCTCTCCTCATCTAGAGGTCTTGCCCCCAATTCCTCTTTCCCAAGACTTA	480
QY	481	TGATTTAGAGTCCACGTGTCCTCCCTCAAAACGGGAGTCCAGTCCCTGCTCAC	540
Db	481	TGATTTAGAGTCCACGTGTCCTCCCTCAAAACGGGAGTCCAGTCCCTGCTCAC	540
QY	541	CAGGCTCAGGATGGGGGTCCCATCCCTGCAATCCAGGCGTCCCGCTGCTGATCA	600
Db	541	CAGGCTCAGGATGGGGGTCCCATCCCTGCAATCCAGGCGTCCCGCTGCTGATCA	600
QY	601	GACACTGACCCCATCTTTGAAACCGACCCCAATCTGCTCGGTGATCAAGGCTCTCGG	660
Db	601	GACACTGACCCCATCTTTGAAACCGACCCCAATCTGCTCGGTGATCAAGGCTCTCGG	660
QY	661	CCAAGGCCAGTCCCTTACAGCTGCGCTGAGATGAGACGCTGGGACTGGGGGGCGCAGAGCT	720
Db	661	CCAAGGCCAGTCCCTTACAGCTGCGCTGAGATGAGACGCTGGGACTGGGGGGCGCAGAGCT	720
QY	721	GGGCTGGGCTGGGCTCCCCAGGCGCTGCTCCCGTCAATCTCTGACAGGTCCACCC	780
Db	721	GGGCTGGGCTGGGCTCCCCAGGCGCTGCTCCCGTCAATCTCTGACAGGTCCACCC	780
QY	781	TGGCCCAAGAGTCAAGCCAGGAATCAATTAACAAGAGGACTGACATGGCGCAGAAAGAG	840
Db	781	TGGCCCAAGAGTCAAGCCAGGAATCAATTAACAAGAGGACTGACATGGCGCAGAAAGAG	840
QY	841	GATGGCCGAGCTGTGCAATGCTGCTCCAGACCCAAAGTGGGAGCTTCACATGCGGGAGCC	900
Db	841	GATGGCCGAGCTGTGCAATGCTGCTCCAGACCCAAAGTGGGAGCTTCACATGCGGGAGCC	900
QY	901	CTGCTACTTCTGACAGCCATCGGGGCGGCAATCTGGGCAATTTGGGCTTTCTCTCAGG	960
Db	901	CTGCTACTTCTGACAGCCATCGGGGCGGCAATCTGGGCAATTTGGGCTTTCTCTCAGG	960
QY	961	AGTACCAAGAGCCGCTGTATCCAGTCAAGTCACTCTGCGAGCGTCCGGCTCATGCTC	1020
Db	961	AGTACCAAGAGCCGCTGTATCCAGTCAAGTCACTCTGCGAGCGTCCGGCTCATGCTC	1020
QY	1021	TTTGAACAAGACGGAAGGAGCGTGGCGGCTGTGTGCTCTCGCGCTCCAAAGCCAGGGTA	1080
Db	1021	TTTGAACAAGACGGAAGGAGCGTGGCGGCTGTGTGCTCTCGCGCTCCAAAGCCAGGGTA	1080
QY	1081	GCGGAGCTCAGCTGCGAGAGATGGGGTTCTCTAGGGGCACTGACCCACTCCGACTGGAC	1140
Db	1081	GCGGAGCTCAGCTGCGAGAGATGGGGTTCTCTAGGGGCACTGACCCACTCCGACTGGAC	1140
QY	1141	GTGCGAACGCGCGGCGCAATGACAGCTCGGGCTTTCTTCTGTGTGACAGAGGGAGGCTG	1200

Accession	Sequence	Length
QY	1021 TTTATCAAGACGGAAGGAGCGTGGCGGCTGTGTGCTTCGCGCTCCAAAGCCAGGGTA	1080
QY	1081 GCCGGACTCAGCTCCAGAGAGATGGGGCTTCCATGAGGCACTGACCCCACTCCGAGCTTGAC	1140
Db	1081 GCCGGACCTCACTCCAGAGAGATGGGGCTTCCATGAGGCACTGACCCCACTCCGAGCTTGAC	1140
QY	1141 GTGCGAACGGCGGGCGCCATGGCACGCTGGGGCTTCTCTGTGTGAGCGAGGGAGGCTG	1200

```

Db      1141  GTGGAAGGCGGGGCGCAATGGACGTCGGCTTCTTGTGTGAGACGAGGGAGGGCTG 1200
Qy      1201  CCCACACCCAGAGGCTGTGGAGTCACTTCGTGTGTGATGTGCCAGAGCCGTTTC 1260
Db      1201  CCCACACCCAGAGGCTGTGGAGTCACTTCGTGTGTGATGTGCCAGAGCCGTTTC 1260
Qy      1261  TTGGCGGCATCTGCGAAGCTGTGGCGGAGAGCTGCCCGGTGAGACCGATCGTGGG 1320
Db      1261  TTGGCGGCATCTGCGAAGCTGTGGCGGAGAGCTGCCCGGTGAGACCGATCGTGGG 1320
Qy      1321  GGGCGGAGACACAGCTTGGGCGGTGGCGGTGGCAAGTCAAGCTTCGCTATGATGAGCA 1380
Db      1321  GGGCGGAGACACAGCTTGGGCGGTGGCGGTGGCAAGTCAAGCTTCGCTATGATGAGCA 1380
Qy      1381  CACCTCTGTGGGGAGTCCCTGCTCCGGGGAGCTGGGTGAGACAGCCGCACTGCTTC 1440
Db      1381  CACCTCTGTGGGGAGTCCCTGCTCCGGGGAGCTGGGTGAGACAGCCGCACTGCTTC 1440
Qy      1441  CCGAGCGGAAACCGGGTCTGTCCGGATGGCGAGTGTTCGGGTGCGGTGGCCAGGAC 1500
Db      1441  CCGAGCGGAAACCGGGTCTGTCCGGATGGCGAGTGTTCGGGTGCGGTGGCCAGGAC 1500
Qy      1501  TCTCCCAACGCTGTGACAGCTGGGGGTGGAGCTGTGTCTACACAGGGGGCTATCTTCC 1560
Db      1501  TCTCCCAACGCTGTGACAGCTGGGGGTGGAGCTGTGTCTACACAGGGGGCTATCTTCC 1560
Qy      1561  TTTGGGAGACCCCAACAGGAGAGAACAGATATGGCTGCTGCTCCACTCTCCAGT 1620
Db      1561  TTTGGGAGACCCCAACAGGAGAGAACAGATATGGCTGCTGCTCCACTCTCCAGT 1620
Qy      1621  CCCCTGCCCCCAAGATATCAAGCTGTGTGCTCCCAAGCTGCGGGGCAAGCCCTG 1680
Db      1621  CCCCTGCCCCCAAGATATCAAGCTGTGTGCTCCCAAGCTGCGGGGCAAGCCCTG 1680
Qy      1681  GTGATGAGCAAGATGTATACCTGTGACGGGCTGGGGCAACAGAGTATATGAGCAAG 1740
Db      1681  GTGATGAGCAAGATGTATACCTGTGACGGGCTGGGGCAACAGAGTATATGAGCAAG 1740
Qy      1741  GCGGGGAGTACTCCAGAGAGCTGAGTCCCATATATCAGCAATGATGTGTGCAATGGCCT 1800
Db      1741  GCGGGGAGTACTCCAGAGAGCTGAGTCCCATATATCAGCAATGATGTGTGCAATGGCCT 1800
Qy      1801  GACTTCTATGAAACCAAGTCAAGGCCCAAGATGTCTGTGCTGCTACCCGAGGGTGGC 1860
Db      1801  GACTTCTATGAAACCAAGTCAAGGCCCAAGATGTCTGTGCTGCTACCCGAGGGTGGC 1860
Qy      1861  ATTGATGCTGCGAGGCGAGAGGAGGCTGCTTGTGTGAGAGAGAGAGTCTCTGG 1920
Db      1861  ATTGATGCTGCGAGGCGAGAGGAGGCTGCTTGTGTGAGAGAGAGAGTCTCTGG 1920
Qy      1921  ACGGCAAGTGGCGGCTGTGTGAGATTTGTGAGTGGGCACTGGCTGTGCTGGCCAG 1980
Db      1921  ACGGCAAGTGGCGGCTGTGTGAGATTTGTGAGTGGGCACTGGCTGTGCTGGCCAG 1980
Qy      1981  AAGCAGAGGCTGTACCAAGTCAAGTCACTTCCGGAGTGAATCTTCCAGGCCATTAAG 2040
Db      1981  AAGCAGAGGCTGTACCAAGTCAAGTCACTTCCGGAGTGAATCTTCCAGGCCATTAAG 2040
Qy      2041  ACTCACTCCGAAGCAGAGGAGTGTGACCCAGCTGTGACGGGTGGCTTCCGCGGCA 2100
Db      2041  ACTCACTCCGAAGCAGAGGAGTGTGACCCAGCTGTGACGGGTGGCTTCCGCGGCA 2100
Qy      2101  GCCTTCAGAGGCGGAGGTATCCGGTGTGTGGATCCAGCTGGAGCCGAGAGTGGAGT 2160
Db      2101  GCCTTCAGAGGCGGAGGTATCCGGTGTGTGGATCCAGCTGGAGCCGAGAGTGGAGT 2160
Qy      2161  TTTTCTTTTGGGCGGGTCCAGAGTCCAGAGAACCTCTCCCTTCAGGGTCTCTCTTC 2220
Db      2161  TTTTCTTTTGGGCGGGTCCAGAGTCCAGAGAACCTCTCCCTTCAGGGTCTCTCTTC 2220
Qy      2221  CACAGTGGCGGCGCACTCAGCCCGGAGACCAACCTCAACCTCTGACCCCAATGT 2280

```

```

Db      2221  CACAGTGGCGGCGCACTCAGCCCGGAGACCAACCTCAACCTCTGACCCCAATGT 2280
Qy      2281  AAATATTTTGTGCTGTCTGTGGAGTCTCTGTGTAGTGTGCTTATGATGGATGCTTTC 2340
Db      2281  AAATATTTTGTGCTGTCTGTGGAGTCTCTGTGTAGTGTGCTTATGATGGATGCTTTC 2340
Qy      2341  AAATATTAAGATGTTTTGATT 2363
Db      2341  AAATATTAAGATGTTTTGATT 2363

RESULT 4
AX511632/c
LOCUS      AX511632      2363 bp      DNA      linear      PAT 27-SEP-2002
DEFINITION Sequence 39 from Patent WO0205705.
ACCESSION  AX511632
VERSION     AX511632.1 GI:23392433
KEYWORDS
SOURCE
ORGANISM   Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS    Mezes, P.S., Rastelli, L., Herrmann, J.L., Macdougall, J.R., Zhong, H.,
            Caeman, S.J., Boldog, F., Shinkets, R.A., Gorman, L., Crasta, O.R.,
            Mysore, K.K., Folkerts, O., Martin, G.B., Eilen, A., Spaderna, S.K.,
            Verne, C.A., Bergh, C., Szytek, K.A., Dippo, V.A., Zernhusen, B.D.,
            Peyman, J.A., Ellerman, K., Stone, D.J., Grose, W.M., Alsdbrook, J.P.,
            Lepley, D.M., Rieger, D.K., Burgess, C.E. and Edinger, S.
            Proteins and nucleic acids encoding same
            Patent: WO 0205705/A 39 18-JUL-2002;
            Curagen Corporation (US)
FEATURES
            Location/Qualifiers
            source
            1..2363
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Query Match      100.0%; Score 2363; DB 6; Length 2363;
Beet Local Similarity 100.0%; Pred. No. 0;
Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  TCGAGCCCGCTTTTCCAGGAGACCTTACCTGAGGGCCCAAGTGTGAGGACCTTGGCTTAC 60
Db      2363  TCGAGCCCGCTTTTCCAGGAGACCTTACCTGAGGGCCCAAGTGTGAGGACCTTGGCTTAC 2304
Qy      61  AGGCCCAAGCGACCGGCTGCTGAGGCGGCGGCTGCTGAGGCGGCGGCGGCGGCGGCGG 120
Db      2303  AGGCCCAAGCGACCGGCTTCTGCTTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2244
Qy      121  TGCCAGGCTGTGAGTGAACCCGACCCGAGCTTACCTGAGGCTCGGCCCACTGTC 180
Db      2243  TGCCAGGCGTGTGAGTGAACCCGAGCCGAGCTTACCTGAGGCTCGGCCCACTGTC 2184
Qy      181  TGAACCCAGGGTGAAGACAAAGGCCCCCAAGTCTCAAGTTCCAGCCCTGAGAGAGGGG 240
Db      2183  TGAACCCAGGGTGAAGACAAAGGCCCCCAAGTCTCAAGTTCCAGCCCTGAGAGAGGGG 2124
Qy      241  TTCCCTATCCCCCAACCGAGCTTAATAGCCCACTCTTAATAGAGGGGTTCTTGGGAGC 300
Db      2123  TTCCCTATCCCCCAACCGAGCTTAATAGCCCACTCTTAATAGAGGGGTTCTTGGGAGC 2064
Qy      301  TGAAGAGGGGCACTATGAGCTTCCCCAGAGCACTTATGATGTTCTGTCTGCTTCTT 360
Db      2063  TGAAGAGGGGCACTATGAGCTTCCCCAGAGCACTTATGATGTTCTGTCTGCTTCTT 2004
Qy      361  CAGACTAGCCGTTTGAACCCCAAGTCTTCTTCTCCAGAGACCCAGAGATTGAGCCCTCAG 420
Db      2003  CAGACTAGCCGTTTGAACCCCAAGTCTTCTTCTCCAGAGACCCAGAGATTGAGCCCTCAG 1944
Qy      421  GCGCCTCTCTCTCATCTAGTGGAGTCTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480

```

Db 1943 GCCCTCCCTCCCTACTACTAGAGAGTCTGGGCCCCCAAAATTCCTCTTTCCCAAGACTTA 1884
Qy 481 TGAATTTAGAGTCTCAGTGTCTCTCTCCCTCAAAACCGGGATCTCAGTCTCCCTGTCTCAC 540
Db 1883 TGAATTTAGAGTCTCAGTGTCTCTCTCCCTCAAAACCGGGATCTCAGTCTCCCTGTCTCAC 1824
Qy 541 CAGGCTCAGAGTGGGGGTCCCAATCCCTGCAAAATCCAGAGGTCCCGGCTGTGTGCA 600
Db 1823 CAGGCTCAGAGTGGGGGTCCCAATCCCTGCAAAATCCAGAGGTCCCGGCTGTGTGCA 1764
Qy 601 GACACTGACCCCATCTTTGAAACCAAGCCAAATCTGCGTGTGATCAGCGGTGTCTGG 660
Db 1763 GACACTGACCCCATCTTTGAAACCAAGCCAAATCTGCGTGTGATCAGCGGTGTCTGG 1704
Qy 661 CCAAGGCCCATCTCCCTAAGCTGTGCTGGATGAGACGCTGGAGACTGGGGGGCCAGAGCT 720
Db 1703 CCAAGGCCCATCTCCCTAAGCTGTGCTGGATGAGACGCTGGAGACTGGGGGGCCAGAGCT 1644
Qy 721 GGGCTGGGCTGGGCTCCCGCAGGCTCTGCTCCCGTCCATCTCTCAAGAGTCCACCC 780
Db 1643 GGGCTGGGCTGGGCTCCCGCAGGCTCTGCTCCCGTCCATCTCTCAAGAGTCCACCC 1584
Qy 781 TGGCCCAAGAGGTGACCCAGGGAATCATTAACAAGGCACTGACATGCGCAGAGAGAG 840
Db 1583 TGGCCCAAGAGGTGACCCAGGGAATCATTAACAAGGCACTGACATGCGCAGAGAGAG 1524
Qy 841 GGTGGCCGGAATGTGCTGCTCCAGACCAAGGTGGAGCTTCACATGCGGGAGACC 900
Db 1523 GGTGGCCGGAATGTGCTGCTCCAGACCAAGGTGGAGCTTCACATGCGGGAGACC 1464
Qy 901 CTGCTACTTCTGACAGCCATCGGGGCGGCACTCTGAGCCATTGTGTGCTCTCTCAGG 960
Db 1463 CTGCTACTTCTGACAGCCATCGGGGCGGCACTCTGAGCCATTGTGTGCTCTCTCAGG 1404
Qy 961 AGTACCAAGAGCGCGTGTATCCAGTGCAGTCACTCTGCGGAGCGTGGCTCATGCTC 1020
Db 1403 AGTACCAAGAGCGCGTGTATCCAGTGCAGTCACTCTGCGGAGCGTGGCTCATGCTC 1344
Qy 1021 TTTTACAAAGAGGGAAGGAGCGTGGCGGTGTGTGCTCTGCGCTCCAAAGCCAGAGTA 1080
Db 1343 TTTTACAAAGAGGGAAGGAGCGTGGCGGTGTGTGCTCTGCGCTCCAAAGCCAGAGTA 1284
Qy 1081 GCCGAGCTCAGCTCGAGAGATGGGCTTCTCAGGGGCACTGACCCACTCCGACTGGAC 1140
Db 1283 GCCGAGCTCAGCTCGAGAGATGGGCTTCTCAGGGGCACTGACCCACTCCGACTGGAC 1224
Qy 1141 GTGCGAAACGGGCGGCCCAATGCAAGTGGGCTTCTTGTGTGTGAGCAGAGGAGGCTG 1200
Db 1223 GTGCGAAACGGGCGGCCCAATGCAAGTGGGCTTCTTGTGTGTGAGCAGAGGAGGCTG 1164
Qy 1201 CCCCACACCAAGAGGCTGTGGAGGTGATCTCGGTGTGTGATGCCCCAGAGGCGTTTC 1260
Db 1163 CCCCACACCAAGAGGCTGTGGAGGTGATCTCGGTGTGTGATGCCCCAGAGGCGTTTC 1104
Qy 1261 TTGGCCGCACTCTGCAAGACTGTGTGCGCAGAGAGCTGCCGTGTGACCCGATGTGGGA 1320
Db 1103 TTGGCCGCACTCTGCAAGACTGTGTGCGCAGAGAGCTGCCGTGTGACCCGATGTGGGA 1044
Qy 1321 GGGCGGGAACACAGCTTGGGCGGCTGGGCGTGTGCAAGTCAAGCTTCCGTATGATGAGCA 1380
Db 1043 GGGCGGGAACACAGCTTGGGCGGCTGGGCGTGTGCAAGTCAAGCTTCCGTATGATGAGCA 984
Qy 1381 CACCTGTGTGGGGATCCCTGTCTCCGGGGACTGGGGTGTGAGAGCCGCCCATGTCTTC 1440
Db 983 CACCTGTGTGGGGATCCCTGTCTCCGGGGACTGGGGTGTGAGAGCCGCCCATGTCTTC 924
Qy 1441 CCGGAGGGAACCGGGTCTGTCCCGATGCGAGTGTGCTTGGCGGTGTGCGCCAGAGCC 1500
Db 923 CCGGAGGGAACCGGGTCTGTCCCGATGCGAGTGTGCTTGGCGGTGTGCGCCAGAGCC 864
Qy 1501 TCTCCCAACGGTGTGACAGTGGGGGTGAGGCTGTGTCTACCAACGGGGGTATCTTCCC 1560
Db 863 TCTCCCAACGGTGTGACAGTGGGGGTGAGGCTGTGTCTACCAACGGGGGTATCTTCCC 804

Qy 1561 TTTTGGGAGCCCCCAACAGGAGGAGAGAGAGATATGTCCTGGTCCACTCTCCAGT 1620
Db 803 TTTTGGGAGCCCCCAACAGGAGGAGAGAGAGATATGTCCTGGTCCACTCTCCAGT 744
Qy 1621 CCCCTGCCCCCTCAAGAAATCATCAGCTGTGTGTCTCCAGCTGTGCGGACAGCCCTG 1680
Db 743 CCCCTGCCCCCTCAAGAAATCATCAGCTGTGTGTCTCCAGCTGTGCGGACAGCCCTG 684
Qy 1681 GTGATGGCAAGATCTGTACCGTGAACGGGCTGGGGCAACAGCAGTACTATGCGCAACAG 1740
Db 683 GTGATGGCAAGATCTGTACCGTGAACGGGCTGGGGCAACAGCAGTACTATGCGCAACAG 624
Qy 1741 GCGGGGTACTCCAGAGAGCTCGAGTCCCATATACAGATGATGTCTGCAATGGGCT 1800
Db 623 GCGGGGTACTCCAGAGAGCTCGAGTCCCATATACAGATGATGTCTGCAATGGGCT 564
Qy 1801 GACTTCTATGAAAACAGATCAAGCCCAAGATGTTCTGTGTGCTGATCCCGAGAGTGGC 1860
Db 563 GACTTCTATGAAAACAGATCAAGCCCAAGATGTTCTGTGTGCTGATCCCGAGAGTGGC 504
Qy 1861 ATTGATGCTGTCCAGGGGCAACAGGCTGTCTTTGTGTGTGAGGACAGCATCTCTCGG 1920
Db 503 ATTGATGCTGTCCAGGGGCAACAGGCTGTCTTTGTGTGTGAGGACAGCATCTCTCGG 444
Qy 1921 ACGCCAGCTTGGGCGGTGTGTGGATGTGAGTTGGGGGCACTGGCTGTGCGCCAG 1980
Db 443 ACGCCAGCTTGGGCGGTGTGTGGATGTGAGTTGGGGGCACTGGCTGTGCGCCAG 384
Qy 1981 AAGCCAGCGCTTACACCAAGTCAAGTCTTCGAGAGTGAATCTTCAAGCCATTAAG 2040
Db 383 AAGCCAGCGCTTACACCAAGTCAAGTCTTCGAGAGTGAATCTTCAAGCCATTAAG 324
Qy 2041 ACTGACTCCGAAGCGAGCGCATGTGTACCCAGCTGTGACCGGTGTCTCTGTGCGCA 2100
Db 323 ACTGACTCCGAAGCGAGCGCATGTGTACCCAGCTGTGACCGGTGTCTCTGTGCGCA 264
Qy 2101 GCCTCAGAGGCGGAGGTATCCCGGTGTGTGGATGCCAGCTGGGCGAGAGATGGAGCT 2160
Db 263 GCCTCAGAGGCGGAGGTATCCCGGTGTGTGGATGCCAGCTGGGCGAGAGATGGAGCT 204
Qy 2161 TTTTCTTCTTGTGGGCGCGGTCAACAGTGCACAGGACACCTCTCCAGAGGTCTCTTTC 2220
Db 203 TTTTCTTCTTGTGGGCGCGGTCAACAGTGCACAGGACACCTCTCCAGAGGTCTCTTTC 144
Qy 2221 CACAGTGGCGGGGCCCATCAGCCCGAGAGACCAACCACTCACCCCTCTGAACCCCATGT 2280
Db 143 CACAGTGGCGGGGCCCATCAGCCCGAGAGACCAACCACTCACCCCTCTGAACCCCATGT 84
Qy 2281 AAATATTGTTCTGTGTGTGGAGTCTCTGTCTGAGTCCCGTGTATGAGATGCTTTC 2340
Db 83 AAATATTGTTCTGTGTGTGGAGTCTCTGTCTGAGTCCCGTGTATGAGATGCTTTC 24
Qy 2341 AAATTAATTAAGATGTTTGATT 2363
Db 23 AAATTAATTAAGATGTTTGATT 1

RESULT 5
HSHERSH
LOCUS 2363 bp mRNA linear PRI 15-JUL-2003
DEFINITION Human hepatoma mRNA for serine protease hepsin.
ACCESSION X07732.1 GI:32063
VERSION X07732.1 GI:32063
KEYWORDS hepsin; membrane protein; serine protease; zymogen.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2363)
AUTHORS Leytus,S.P., Loeb,K.R., Hagen,F.S., Kurachi,K. and Davie,E.W.
TITLE A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells

JOURNAL Biochemistry 27 (3), 1067-1074 (1988)
 MEDLINE 88209431
 PUBMED 2835076

COMMENT see X07002 for liver hepsin partial cDNA sequence
 the authors combined the sequence from several overlapping Hep G2
 cDNA clones; additional sequences were found in clones HepG2UW17
 and HepG2UW2

FEATURES

Source Location/Qualifiers
 1..2363
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="HepG2UW63, HepG2UW61, HepG2UW17, HepG2UW2"
 /cell_line="Hep G2"
 /tissue_type="liver"
 /clone_11b="lambda gt11"
 1..191
 /product="HepG2UW7 mRNA"
 83
 /note="5' end of HepG2UW19"
 89
 /note="5' end of HepG2UW61"
 98
 /note="5' end of HepG2UW20"
 116
 /note="5' end of HepG2UW17"
 132
 /note="5' end of HepG2UW63 and HepG2UW2"
 192..771
 /note="Insert in HepG2UW17 (compared to HepG2UW7; put.
 unspliced intron)"
 627..771
 /note="Insert in HepG2UW2 (compared to HepG2UW7; put.
 unspliced intron, part.)"
 772..2363
 /product="HepG2UW7 mRNA"
 826..2079
 /codon_start=1
 /product="hepsin"
 /protein_id="CAA30558.1"
 /db_xref="GI:32064"
 /db_xref="GOA:P05981"
 /db_xref="UniProt/Swiss-Prot:P05981"
 /translation="MAQKEGRTVCCSRFVAAVLTAGTLLLTALGAASVAIVALL
 RSDQPLYPVQVSSADALNWFDTKEGTRLLCSSRNARVAGSCBEMGFLALTHS
 ELIDVETAGNLSGFPCVDEGRLEPHOTRLLEIVSCDPRGRFLAALCODGRALTS
 DRIVGRDTSLSGRWPMQVSLRYDGAHLGGSLSGDWYLTAAHCFPERNRVLSRWVF
 AGAVALASPHGLDGLQAVLVHGVLEPRDPSRNSNDIALVHLSPLPLEYIOPV
 CLPAAGQALVDGKICTVYGMKNTQYTGQAGVLDQAVPIIISNDVNCGADPYGQIQIR
 KMFCAGYPEGGIDACQDGGSPVCEBSISHTPRWRLGIVSWGTCALAQKPGVYTK
 VSDPREWIFQAIKTHSEASGMVTL"
 826..1311
 1312..2076
 /product="unnamed"
 2363

ORIGIN

Query Match 100.0%; Score 2363; DB 9; Length 2363;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TCAGAGCCGCTTTCCAGGAGACCTTACTGAGGAGCCACAGGTGAGGAGCCTGAGCTTAC
 60
 1 TCAGAGCCGCTTTCCAGGAGACCTTACTGAGGAGCCACAGGTGAGGAGCCTGAGCTTAC
 60
 61 AGGCCCCAGCCAGCCTGCTGCTCCAGGCGCCCGCTGCTGCGGGGCGACATGCTCC
 120
 61 AGGCCCCAGCCAGCCTGCTGCTCCAGGCGCCCGCTGCTGCGGGGCGACATGCTCC
 120
 61 AGGCCCCAGCCAGCCTGCTGCTCCAGGCGCCCGCTGCTGCGGGGCGACATGCTCC
 120
 121 TGCCAGGCTGAGAGCTGAGCCGACCCGCGACCTTACGAGGCTCGGCCCGACCTTGC
 180
 121 TGCCAGGCTGAGAGCTGAGCCGACCCGCGACCTTACGAGGCTCGGCCCGACCTTGC
 180
 121 TGCCAGGCTGAGAGCTGAGCCGACCCGCGACCTTACGAGGCTCGGCCCGACCTTGC
 180

181 TGAGCCCCAGGGGTAAAGACAAAGGGGCCCCAGACTCAGATTCCAGCCCTGAGGACAGGGG 240
 181 TGAGCCCCAGGGGTAAAGACAAAGGGGCCCCAGACTCAGATTCCAGCCCTGAGGACAGGGG 240
 241 TTCCCTCATCCCCCACCAGCCTTAATGACCACTCTTAATAGAGGGGTTCTGGGAGCC 300
 241 TTCCCTCATCCCCCACCAGCCTTAATGACCACTCTTAATAGAGGGGTTCTGGGAGCC 300
 241 TTCCCTCATCCCCCACCAGCCTTAATGACCACTCTTAATAGAGGGGTTCTGGGAGCC 300
 301 TGAAGAGGGGCACTATGACGCTCTCCCAAGACCTTATAGTGTCTGTCTGCTCTTCTT 360
 301 TGAAGAGGGGCACTATGACGCTCTCCCAAGACCTTATAGTGTCTGTCTGCTCTTCTT 360
 301 TGAAGAGGGGCACTATGACGCTCTCCCAAGACCTTATAGTGTCTGTCTGCTCTTCTT 360
 361 CAGACTAGGCGTTGAGACCCAGTCCTTCTCCAGACCCAGAGGTTCCAGGCTCAG 420
 361 CAGACTAGGCGTTGAGACCCAGTCCTTCTCCAGACCCAGAGGTTCCAGGCTCAG 420
 361 CAGACTAGGCGTTGAGACCCAGTCCTTCTCCAGACCCAGAGGTTCCAGGCTCAG 420
 421 GCCCTCTCTCCCTCATATAGGAGGCTCTGGCCCCCAATTCCTCTTCCCAAGACTTA 480
 421 GCCCTCTCTCCCTCATATAGGAGGCTCTGGCCCCCAATTCCTCTTCCCAAGACTTA 480
 421 GCCCTCTCTCCCTCATATAGGAGGCTCTGGCCCCCAATTCCTCTTCCCAAGACTTA 480
 481 TGATTTAGAGTCTCAGCTGTCTCTCTCAAAACGGAGATCTGATGCCCTGCTCCAC 540
 481 TGATTTAGAGTCTCAGCTGTCTCTCTCAAAACGGAGATCTGATGCCCTGCTCCAC 540
 481 TGATTTAGAGTCTCAGCTGTCTCTCTCAAAACGGAGATCTGATGCCCTGCTCCAC 540
 541 CAGGCTCAGGATGAGGGGTCCCAATCCCTGCAATTCAGAGGTCCTCCCGCTGCTGCA 600
 541 CAGGCTCAGGATGAGGGGTCCCAATCCCTGCAATTCAGAGGTCCTCCCGCTGCTGCA 600
 541 CAGGCTCAGGATGAGGGGTCCCAATCCCTGCAATTCAGAGGTCCTCCCGCTGCTGCA 600
 601 GACACTAGACCCCATCTTGAACCCAGGCGCAATCTGCTCGTATCAAGGCGTCTGCG 660
 601 GACACTAGACCCCATCTTGAACCCAGGCGCAATCTGCTCGTATCAAGGCGTCTGCG 660
 601 GACACTAGACCCCATCTTGAACCCAGGCGCAATCTGCTCGTATCAAGGCGTCTGCG 660
 661 CCAAGGCGCAGTCCCTTACAGCTGCTGATGAGCGCTTGAGACTGAGGAGCTGAGGAGACT 720
 661 CCAAGGCGCAGTCCCTTACAGCTGCTGATGAGCGCTTGAGACTGAGGAGCTGAGGAGACT 720
 661 CCAAGGCGCAGTCCCTTACAGCTGCTGATGAGCGCTTGAGACTGAGGAGCTGAGGAGACT 720
 721 GGGCTGAGGCTGGGCTCCCCCAGAGGCTGCTCCCGCTCAATCTCTACAGGTCACACC 780
 721 GGGCTGAGGCTGGGCTCCCCCAGAGGCTGCTCCCGCTCAATCTCTACAGGTCACACC 780
 721 GGGCTGAGGCTGGGCTCCCCCAGAGGCTGCTCCCGCTCAATCTCTACAGGTCACACC 780
 781 TGCCCAAGAGGTGAGGAGGAGTCAATTAACAAGAGGAGTGAATGAGCAGCAAGAGAG 840
 781 TGCCCAAGAGGTGAGGAGGAGTCAATTAACAAGAGGAGTGAATGAGCAGCAAGAGAG 840
 781 TGCCCAAGAGGTGAGGAGGAGTCAATTAACAAGAGGAGTGAATGAGCAGCAAGAGAG 840
 841 GGTGGCGGAGCTGTGCGAGCTGCTCCAGACCCAGGTGAGCTTCACTGCGGGAGCC 900
 841 GGTGGCGGAGCTGTGCGAGCTGCTCCAGACCCAGGTGAGCTTCACTGCGGGAGCC 900
 841 GGTGGCGGAGCTGTGCGAGCTGCTCCAGACCCAGGTGAGCTTCACTGCGGGAGCC 900
 901 CTGCTACTTCTGACAGGCAATCGGGGCGGCAATCTGAGGCAATGAGGCTTCTCTCAGG 960
 901 CTGCTACTTCTGACAGGCAATCGGGGCGGCAATCTGAGGCAATGAGGCTTCTCTCAGG 960
 901 CTGCTACTTCTGACAGGCAATCGGGGCGGCAATCTGAGGCAATGAGGCTTCTCTCAGG 960
 961 AGTACAGGAGGCGGCTGTACCCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1020
 961 AGTACAGGAGGCGGCTGTACCCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1020
 961 AGTACAGGAGGCGGCTGTACCCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1020
 961 AGTACAGGAGGCGGCTGTACCCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1020
 1021 TTGACAAAGACGAGAGGAGCTGTGCGGCTGTGCTCTTCTGCGCTCCAAAGGAGGTA 1080
 1021 TTGACAAAGACGAGAGGAGCTGTGCGGCTGTGCTCTTCTGCGCTCCAAAGGAGGTA 1080
 1021 TTGACAAAGACGAGAGGAGCTGTGCGGCTGTGCTCTTCTGCGCTCCAAAGGAGGTA 1080
 1081 GCGGAGTCAAGCTGAGAGGAGTGGGCTTCTTCAAGGCACTGACCCCACTCCAGAGTGA 1140
 1081 GCGGAGTCAAGCTGAGAGGAGTGGGCTTCTTCAAGGCACTGACCCCACTCCAGAGTGA 1140
 1081 GCGGAGTCAAGCTGAGAGGAGTGGGCTTCTTCAAGGCACTGACCCCACTCCAGAGTGA 1140
 1141 GTGCGAAGCGGCGGCGCAATGCGAGCTCGGGCTTCTTCTGTGTGAGACGAGGGAGGCTG 1200
 1141 GTGCGAAGCGGCGGCGCAATGCGAGCTCGGGCTTCTTCTGTGTGAGACGAGGGAGGCTG 1200
 1141 GTGCGAAGCGGCGGCGCAATGCGAGCTCGGGCTTCTTCTGTGTGAGACGAGGGAGGCTG 1200
 1201 CCCCACACCCAGAGGCTGTGAGAGTATCTCGTGTGTATGAGCCCAAGAGGCGGTTTC 1260
 1201 CCCCACACCCAGAGGCTGTGAGAGTATCTCGTGTGTATGAGCCCAAGAGGCGGTTTC 1260
 1201 CCCCACACCCAGAGGCTGTGAGAGTATCTCGTGTGTATGAGCCCAAGAGGCGGTTTC 1260
 1261 TTGGCCGCGCATCTGCGCAAGACTGTGGCGGACGAGAGCTGCCCGTGTGAGCCGATGTGGGA 1320

Db 1261 TTGGCCGCGCATCTGCGAAGACGTGGCCGCGAAGAGCGCCCGTGGACCGCATGCTGGGA 1320
Qy 1321 GGGCGGGGACACCGAGCTTGGGCGCGGTGGCGGCTGGCAAGTCAAGCTTCCGTATGATGAGACA 1380
Db 1321 GGGCGGGGACACCGAGCTTGGGCGCGGTGGCGGCTGGCAAGTCAAGCTTCCGTATGATGAGACA 1380
Qy 1381 CACCTCTGTGGGGGATCCCTGCTGCTCCGGGAGCTGGGGCTGAGACAGCCGCCACGCTGTC 1440
Db 1381 CACCTCTGTGGGGGATCCCTGCTGCTCCGGGAGCTGGGGCTGAGACAGCCGCCACGCTGTC 1440
Qy 1441 CCGAGCGGAAACCGGGTCTGTCCCGATGCGAGTGTGGCCGGTCCGGTGGCCCGACAGCC 1500
Db 1441 CCGAGCGGAAACCGGGTCTGTCCCGATGCGAGTGTGGCCGGTCCGGTGGCCCGACAGCC 1500
Qy 1501 TCTCCCGACCGGTCTGCACTGGGGGTGCAAGCTGTGTCTACACAGGGGGCTATCTTCC 1560
Db 1501 TCTCCCGACCGGTCTGCACTGGGGGTGCAAGCTGTGTCTACACAGGGGGCTATCTTCC 1560
Qy 1561 TTTGGGGAGCCCAACAGCGAGAGAGAACAGAACTATTTGGCCCTGGTCCACTCTCCAGT 1620
Db 1561 TTTGGGGAGCCCAACAGCGAGAGAGAACAGAACTATTTGGCCCTGGTCCACTCTCCAGT 1620
Qy 1621 CCCCTGCCCCACAGAAATACATCAGCTGTGTGCTCCAGCTGCGCGCAGGCCCCCTG 1680
Db 1621 CCCCTGCCCCACAGAAATACATCAGCTGTGTGCTCCAGCTGCGCGCAGGCCCCCTG 1680
Qy 1681 GTGATGCGCAAGATCTGTACCGGTGACGGGCTGGGGCAACAGCGAGTACTATGGCCAAAG 1740
Db 1681 GTGATGCGCAAGATCTGTACCGGTGACGGGCTGGGGCAACAGCGAGTACTATGGCCAAAG 1740
Qy 1741 GCGCGGGGATCTCCAGAGAGCTCGAGTCCCATTAATCAGCAATGATGTGTGCAATGAGGCT 1800
Db 1741 GCGCGGGGATCTCCAGAGAGCTCGAGTCCCATTAATCAGCAATGATGTGTGCAATGAGGCT 1800
Qy 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGGGTGGC 1860
Db 1801 GACTTCTATGAGAAACAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCGAGGGTGGC 1860
Qy 1861 ATTATGCTGTGCGCAGGGGAGACAGGGGTGTCCCTTTGTGTGAGAGACAGATCTCTCGG 1920
Db 1861 ATTATGCTGTGCGCAGGGGAGACAGGGGTGTCCCTTTGTGTGAGAGACAGATCTCTCGG 1920
Qy 1921 ACGCCAGGTTGGCGGCTGTGTGCAATTTGTGAGTTGGGGCACTGGCTGTGGCCCTGAGCCAG 1980
Db 1921 ACGCCAGGTTGGCGGCTGTGTGCAATTTGTGAGTTGGGGCACTGGCTGTGGCCCTGAGCCAG 1980
Qy 1981 AAGCCAGGCGTCTACCAACCAAGTCAAGTCTTCCGGAGTGGATCTTCCAGGCCATTAAG 2040
Db 1981 AAGCCAGGCGTCTACCAACCAAGTCAAGTCTTCCGGAGTGGATCTTCCAGGCCATTAAG 2040
Qy 2041 ACTACCTCCGAAGCGAGGGATGTGACCCAGCTCTGACCGGTGTCTTCCGTGGGCA 2100
Db 2041 ACTACCTCCGAAGCGAGGGATGTGACCCAGCTCTGACCGGTGTCTTCCGTGGGCA 2100
Qy 2101 GCTCCAGAGGCGCAGAGTATCCCGGTGTGAGTTCACGCTGGGCGAGGATGGAGCT 2160
Db 2101 GCTCCAGAGGCGCAGAGTATCCCGGTGTGAGTTCACGCTGGGCGAGGATGGAGCT 2160
Qy 2161 TTTTCTTGTGGGCGCGGTCCACAGGTCCAGGACACCCCTCCCTCAAGGATCTCTCTTC 2220
Db 2161 TTTTCTTGTGGGCGCGGTCCACAGGTCCAGGACACCCCTCCCTCAAGGATCTCTCTTC 2220
Qy 2221 CACAGTGGCGGCGCCTCAAGCCCCGAGACCAACCAACTCAACCTCTCTGACCCCATGT 2280
Db 2221 CACAGTGGCGGCGCCTCAAGCCCCGAGACCAACCAACTCAACCTCTCTGACCCCATGT 2280
Qy 2281 AAATATTGTTCTGCTGTCTGGGACTCTGTCTAGAGTGCCCTGTAGTGGATGCTCTTT 2340
Db 2281 AAATATTGTTCTGCTGTCTGGGACTCTGTCTAGAGTGCCCTGTAGTGGATGCTCTTT 2340
Qy 2341 AAATTAATTAAGATGTTTGAAT 2363
Db 2341 AAATTAATTAAGATGTTTGAAT 2363

Db 2341 AAATTAATTAAGATGTTTGAAT 2363
RESULT 6
CQ718142 1783 bp DNA linear PAT 03-FEB-2004
LOCUS Sequence 4076 from Patent WO02068579.
DEFINITION CQ718142
ACCESSION CQ718142
VERSION CQ718142.1 GI:42278999
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kites, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 4076 06-SBP-2002;
FEATURES
location/Qualifiers
source
1..1783
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 67.4%; Score 1593.6; DB 6; Length 1783;
Best Local Similarity 99.1%; Pred. No. 2.4e-295;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
Qy 736 CCCCAGAGCCCTGCTCCCGCTCCATCTCTCAAGAGTCCACCTGCGCCAGAGGTCA 795
Db 157 CTTGAGGCTTCGCCCCACCTGTGAGACCCCA-GGGTCCACCTGCGCCAGAGGTCA 215
Qy 796 GCCAGGAAATATTAAACAGAGGAGTGAATGAGCGCAGAGAGAGGAGTGGCCGAGCTGTG 855
Db 216 GCCAGGAAATATTAAACAGAGGAGTGAATGAGCGCAGAGAGAGGAGTGGCCGAGCTGTG 275
Qy 856 CCATGCTGTCCAGACCCCAAGGTGAGCTGTCTACTGCGGGACCTGTCTACTTGTACA 915
Db 276 CCATGCTGTCCAGACCCCAAGGTGAGCTGTCTACTGCGGGACCTGTCTACTTGTACA 335
Qy 916 GCCATGGGGGCGGATCTGGGCGCATTTGGGCTTTCTCTCAAGAGTGAACAGAGCCG 975
Db 336 GCCATGGGGGCGGATCTGGGCGCATTTGGGCTTTCTCTCAAGAGTGAACAGAGCCG 395
Qy 976 CTGTACCCAGTGAAGTCAAGCTCTGCGGAGCTTCGCTCATGTCTTTGAACAGAGAA 1035
Db 396 CTGTACCCAGTGAAGTCAAGCTCTGCGGAGCTTCGCTCATGTCTTTGAACAGAGAA 455
Qy 1036 GGGAGCTGGCGGCTGTGTCTCTCTGCGCTCCAGCGCCAGGGTATGCCGACTCAGCTGC 1095
Db 456 GGGAGCTGGCGGCTGTGTCTCTCTGCGCTCCAGCGCCAGGGTATGCCGACTCAGCTGC 515
Qy 1096 GAGAGATGGGCTTCTCAAGGAGTGAACCCACATCCGAGCTGAGCGTGGACGCGAGCG 1155
Db 516 GAGAGATGGGCTTCTCAAGGAGTGAACCCACATCCGAGCTGAGCGTGGACGCGAGCG 575
Qy 1156 GCCAATGGCAGTCTGGGCTTCTCTGTGTGAGAGAGGGAGGTGCCCAACCCAGAGG 1215
Db 576 GCCAATGGCAGTCTGGGCTTCTCTGTGTGAGAGAGGGAGGTGCCCAACCCAGAGG 635
Qy 1216 CTGTGAGAGTCAATCTCGTGTGATTTGCCCAAGAGCGGTTCTTTGGCCCATCTGC 1275
Db 636 CTGTGAGAGTCAATCTCGTGTGATTTGCCCAAGAGCGGTTCTTTGGCCCATCTGC 695
Qy 1276 CAAGACTGTGGCGCGAGAGAGTGGCCCGTGGACCGCATGTGGGAGGCGGAGACACGAGC 1335
Db 696 CAAGACTGTGGCGCGAGAGAGTGGCCCGTGGACCGCATGTGGGAGGCGGAGACACGAGC 755
Qy 1336 TTGGCGCGGTGGCGGTGGCAAGTCAAGCTTCCGTATGATGAGACACACTCTGTGGGGGA 1395
Db 1336 TTGGCGCGGTGGCGGTGGCAAGTCAAGCTTCCGTATGATGAGACACACTCTGTGGGGGA 1395

Db 756 TTGGGCGCGTGGCGTGGCAATGCTGCTTGGCTATGATGAGACACACTCTGTGGGGA 815
Qy 1396 TCCCTGCTCTCCGGGGAAGTGGGTGTGACAGCCGCCCACTGCTTCCGGAGCGGAACCGG 1455
Db 816 TCCCTGCTCTCCGGGGAAGTGGGTGTGACAGCCGCCCACTGCTTCCGGAGCGGAACCGG 875
Qy 1456 GTCCGTGCTCCGATGGCGAGTGTGGTGGCGCGTGGCGCGAGCGCTTCCCAAGCTGTG 1515
Db 876 GTCCGTGCTCCGATGGCGAGTGTGGTGGCGCGTGGCGCGAGCGCTTCCCAAGCTGTG 935
Qy 1516 CAGCTGGGGGTGACAGCTGTGTGTCTACACAGCGGGGTATCTTCTTCCGGAGCGGAAC 1575
Db 936 CAGCTGGGGGTGACAGCTGTGTGTCTACACAGCGGGGTATCTTCTTCCGGAGCGGAAC 995
Qy 1576 AGCGAGGAAACAGCAAGATATGCTGCTGCTCACTCTCAAGTCCCTGCTCCCTCA 1635
Db 996 AGCGAGGAAACAGCAAGATATGCTGCTGCTCACTCTCAAGTCCCTGCTCCCTCA 1055
Qy 1636 GAATATCATCCAGCTGTGTGTCTCCAGCTGCGGCGCGAGCGCTGTGTGTGTGTGTGTG 1695
Db 1056 GAATATCATCCAGCTGTGTGTCTCCAGCTGCGGCGCGAGCGCTGTGTGTGTGTGTGTG 1115
Qy 1696 TGTACCGTGAAGCGGCTGGGGGCAACAGCAATATGAGCAAGCGCGGGGTACTCCAG 1755
Db 1116 TGTACCGTGAAGCGGCTGGGGGCAACAGCAATATGAGCAAGCGCGGGGTACTCCAG 1175
Qy 1756 GAGGCTGAGTCCCATATATGAGCAAGTATGCTGTGCAATGAGCGCTGATCTTATAGAA 1815
Db 1176 GAGGCTGAGTCCCATATATGAGCAAGTATGCTGTGCAATGAGCGCTGATCTTATAGAA 1235
Qy 1816 CAGATCAAGCGGCAAGATGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1875
Db 1236 CAGATCAAGCGGCAAGATGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1295
Qy 1876 GCGCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1935
Db 1296 GCGCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1355
Qy 1936 CTGTGTGCAATG 1995
Db 1356 CTGTGTGCAATG 1415
Qy 1996 ACCAAAGTCAAGTG 2055
Db 1416 ACCAAAGTCAAGTG 1475
Qy 2056 AGCGGCAATG 2115
Db 1476 AGCGGCAATG 1535
Qy 2116 GGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2175
Db 1536 GGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1595
Qy 2176 CGGTTCACAGGTCCAGAGCAACCTCTCCAGGGTCTCTCTTCCAGTGTGGCGGCGC 2235
Db 1596 CGGTTCACAGGTCCAGAGCAACCTCTCCAGGGTCTCTCTTCCAGTGTGGCGGCGC 1655
Qy 2236 ACTAGAGCGGAGCAACCAAGCTCAACCTCTGAGAGCGGAGCTGAAATATGTTCTGCT 2295
Db 1656 ACTAGAGCGGAGCAACCAAGCTCAACCTCTGAGAGCGGAGCTGAAATATGTTCTGCT 1715
Qy 2296 GTCTGGAAGTCTCTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 2355
Db 1716 GTCTGGAAGTCTCTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1775
Qy 2356 TTTTGATT 2363
Db 1776 TTTTGATT 1783

RESULT 7

AR280046
LOCUS AR280046 1783 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 188 from patent US 6518028.
ACCESSION AR280046
VERSION AR280046.1 GI:29715235
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
AUTHORS O'Brien, T.L.
TITLE Methods for the early diagnosis of ovarian and prostate cancer
JOURNAL Patent: US 6518028-A 188 11-FEB-2003;
FEATURES
Source Location/Qualifiers
1. 1783
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 67.4% Score=1593.67 DB 6; Length 1783;
Best Local Similarity 99.1% Pred. No. 2.4e-25;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

Qy 736 CCCCCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 795
Db 157 CTTGAGGCTTCCGCCCCCACTGCTGAGACCCA-GGGTCCACCTGGCCAGAGGTCA 215
Qy 796 GCCAGGATCATTTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 855
Db 216 GCCAGGATCATTTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 275
Qy 856 CCATGCTGCTCAGACCCAGAGGTGAGCTTCACTGCGGGAGACCTGTCTACTGACA 915
Db 276 CCATGCTGCTCAGACCCAGAGGTGAGCTTCACTGCGGGAGACCTGTCTACTGACA 335
Qy 916 GCCATGCGGGGCGGATCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 975
Db 336 GCCATGCGGGGCGGATCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 395
Qy 976 CTGTACCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1035
Db 396 CTGTACCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 455
Qy 1036 GGGAGCTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1095
Db 456 GGGAGCTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 515
Qy 1096 GAGGAGATGGGCTTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1155
Db 516 GAGGAGATGGGCTTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 575
Qy 1156 GCCAATGAGCAAGTGGGCTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1215
Db 576 GCCAATGAGCAAGTGGGCTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 635
Qy 1216 CTGTGAGAGTCAATCTCGTG 1275
Db 636 CTGTGAGAGTCAATCTCGTG 695
Qy 1276 CAAGACGTGGCGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1335
Db 696 CAAGACGTGGCGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 755
Qy 1336 TTGGGCGGAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1395
Db 756 TTGGGCGGAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 815
Qy 1396 TCCCTGCTCTCCGGGGAAGTGGGTGTGACAGCGGCCCACTGCTGTGGGGGA 1455
Db 816 TCCCTGCTCTCCGGGGAAGTGGGTGTGACAGCGGCCCACTGCTGTGGGGGA 875
Qy 1456 GTCCGTGCTCCGATGGCGAGTGTGGTGGCGCGTGGCGCGAGCGCTTCCCAAGCTGTG 1515


```

Db      756 TTGGGCGGGTGGCCCTGGGAATGACGCTTCCCTATGATGAGACACACTCTGTGGGGG 815
Qy      1396 TCCCTGCTCTCCGGGAGACTGGGTGCTGACAGCCGCCCACTGTTCCCGAGGGAACCGG 1455
Db      816 TCCCTGCTCTCCGGGAGACTGGGTGCTGACAGCCGCCCACTGTTCCCGAGGGAACCGG 875
Qy      1456 GTCCCTGCTCCGATGGCGAGTCTTTTCCGGGTGGCCGAGGCTCTTCCCAAGGCTGTG 1515
Db      876 GTCCCTGCTCCGATGGCGAGTCTTTTCCGGGTGGCCGAGGCTCTTCCCAAGGCTGTG 935
Qy      1516 CAGCTGGGGGAGTGGGTGCTACCAAGGGGGGCTATCTTCCCTTTGAGGAGCCCAAC 1575
Db      936 CAGCTGGGGGAGTGGGTGCTACCAAGGGGGGCTATCTTCCCTTTGAGGAGCCCAAC 995
Qy      1576 AGCGAGAGAACAGACAGATATGCTCCGTGTCACTTCTCCAGTCCCTGCGCTCA 1635
Db      996 AGCGAGAGAACAGACAGATATGCTCCGTGTCACTTCTCCAGTCCCTGCGCTCA 1055
Qy      1636 GAATTCATCCAGCTGTGTGCTCCAGCTGCGCGGCGAGGCTCTGTGGATGGCAAGATC 1695
Db      1056 GAATTCATCCAGCTGTGTGCTCCAGCTGCGCGGCGAGGCTCTGTGGATGGCAAGATC 1115
Qy      1696 TGTACCGTGAAGGCGGCTGGGGCAACAGCGAGTACTATGAGCAAGAGGCGGCTTCCAG 1755
Db      1116 TGTACCGTGAAGGCGGCTGGGGCAACAGCGAGTACTATGAGCAAGAGGCGGCTTCCAG 1175
Qy      1756 GAGGCTGAGTCCCATATACGACATGATGTCTGCAATGGCGCTGACTTCTATGAAAC 1815
Db      1176 GAGGCTGAGTCCCATATACGACATGATGTCTGCAATGGCGCTGACTTCTATGAAAC 1235
Qy      1816 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGGTGGCATTTGATGCTCCAG 1875
Db      1236 CAGATCAAGCCCAAGATGTTCTGTGCTGCTACCCCGAGGGTGGCATTTGATGCTCCAG 1295
Qy      1876 GAGCGACAGGAGTGGCTGCTGTGTGATGAGGACAGACTCTCGGAGCGCAAGTTGGGAG 1935
Db      1296 GAGCGACAGGAGTGGCTGCTGTGTGATGAGGACAGACTCTCGGAGCGCAAGTTGGGAG 1355
Qy      1936 CTGTGTGCAATTTGATGTTGGGGAAGTGGCTGTGTGCTGCGCCAGAGCCAGGCGCTTAC 1995
Db      1356 CTGTGTGCAATTTGATGTTGGGGAAGTGGCTGTGTGCTGCGCCAGAGCCAGGCGCTTAC 1415
Qy      1996 ACCAAGTCAATGATTTCCGGGAGTGGATCTTCCAGGCGCATTAAGACTCTCCGAAGCC 2055
Db      1416 ACCAAGTCAATGATTTCCGGGAGTGGATCTTCCAGGCGCATTAAGACTCTCCGAAGCC 1475
Qy      2056 AGCGGATGATGACCCAGCTGACCGGAGGCTTCTGCTGCGAGGCTCCAGAGGCGCGA 2115
Db      1476 AGCGGATGATGACCCAGCTGACCGGAGGCTTCTGCTGCGAGGCTCCAGAGGCGCGA 1535
Qy      2116 GATGATCCCGGTGATGAGATCCAGCTGAGCGGCGAGAGATGGACGTTTCTTCTTGGGCG 2175
Db      1536 GATGATCCCGGTGATGAGATCCAGCTGAGCGGCGAGAGATGGACGTTTCTTCTTGGGCG 1595
Qy      2176 CCGTTCACAGATCCCAAGACACCTTCTCCAGGAGTCTTCTTCTTCCAGATGGGGGCGC 2235
Db      1596 CCGTTCACAGATCCCAAGACACCTTCTCCAGGAGTCTTCTTCTTCCAGATGGGGGCGC 1655
Qy      2236 ACTGAGCGCGAGACACCAACCTGACCTCTGACCCGCAAGGATGTAATTTGTTCTGCT 2295
Db      1656 ACTGAGCGCGAGACACCAACCTGACCTCTGACCCGCAAGGATGTAATTTGTTCTGCT 1715
Qy      2296 GTCTGGAGCTCTCTGTAGTGGCCCTGATGATGATGGAGTCTTTTAAATAAAGATGG 2355
Db      1716 GTCTGGAGCTCTCTGTAGTGGCCCTGATGATGATGGAGTCTTTTAAATAAAGATGG 1775
Qy      2356 TTTTGATT 2363
Db      1776 TTTTGATT 1783

```

RESULT 9

```

HUMHPSNA
LOCUS HUMHPSNA 1783 bp mRNA linear PRI 11-JUN-1993
DEFINITION Human hepsin mRNA, complete cds.
ACCESSION M18930
VERSION M18930.1 GI:184371
KEYWORDS hepsin; protease; serine protease.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Leytus,S.P., Loeb,K.R., Hagen,F.S., Kurachi,K. and Davie,E.W.
TITLE A novel trypsin-like serine protease (hepsin) with a putative
transmembrane domain expressed by human liver and hepatoma cells
JOURNAL Biochemistry 27 (3), 1067-1074 (1988)
MEDLINE 88209431
PUBMED 2835076
COMMENT Original source text: Human liver cell, cDNA to mRNA, clone
HepG2UM7.
FEATURES
Source location/Qualifiers
1..1783
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1783
/product="hepsin mRNA"
246..1499
/note="hepsin (serine protease) precursor"
/codon_start=1
/protein_id="AAA36013.1"
/db_xref="GI:306886"
/translacion="MAQREGRTVPCSPRYAAALNGLTLLTAIGASNAIYAVLL
RSDPEYPVQVSSADARLWFDTEBTWKLCSRSNARVAGISCMEFLALVHS
ELDVRTAGANGTSQFPCVDBGRLLPHTLLLEIVSVCCPRPRPLAALCODGRRLPV
DRIVGRDTSIGRPMQVSLRYDGLACGSLSGDWLTAHCFPERNRVLSHRV
AGAAVQASPHGLQIGVAVVYHGGTLPPRDNSENNDIALVHLSPLPLEYIOPV
CLPAAGALVNDKICTYTGKNTGYOQAGLPAEDAVPIISNVCCADPYQNDIQP
KMFCAQIPBEGSIDLCQDSGSPVCEISISRTPRMLGIVSMGTGCLAAKPGVYTK
VSDREWIFQAIKTHSEASGWTQL"
246..731
/note="hepsin signal peptide"
732..1496
/product="hepsin"
297..377
/misc_feature
/note="transmembrane domain"
ORIGIN 35 bp upstream of Apai site.
Query Match 67.4%; Score 1593.6; DB 9; Length 1783;
Best Local Similarity 99.1%; Pred. No. 2.4e-295;
Matches 1613; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
Qy 736 CCCCCAGGCCCTGCTCCCTCCATCTCTCACAGTCCCAACCTGGCCGAGAGGTCA 795
Db 157 CTTGAGGCTCTCGGCCCACTGCTGAGACCCCA-GGGTCCCACTGGCCGAGAGGTCA 215
Qy 796 GCCAGGGAATCATTAACAAGGCAATGACATGCGCGAGAAAGAGGTTGCCGACTGTG 855
Db 216 GCCAGGGAATCATTAACAAGGCAATGACATGCGCGAGAAAGAGGTTGCCGACTGTG 275
Qy 856 CCATGCTGCTCACAACCAAGGTGGAGCTCTCACTGGGGGAGCCCTGCTACTTCTGACA 915
Db 276 CCATGCTGCTCACAACCAAGGTGGAGCTCTCACTGGGGGAGCCCTGCTACTTCTGACA 335
Qy 916 GCCATCGGGGGGAGTCTGGGCCATTTGTGGCTGTTCTCTCAGAGTGAACAGAGCGG 975
Db 336 GCCATCGGGGGGAGTCTGGGCCATTTGTGGCTGTTCTCTCAGAGTGAACAGAGCGG 395
Qy 976 CTGTACCAAGTGCAGGTCAAGTCTGCGGAGGCTTGCGCTCATGTTTACAAAGCGAA 1035
Db 396 CTGTACCAAGTGCAGGTCAAGTCTGCGGAGGCTTGCGCTCATGTTTACAAAGCGAA 455
Qy 1036 GGGACGTGGGGGCTGTGTGCTCTCGCGCTCAACCGCAGGGTACCGGACTGAGCTGC 1095

```

Db	456	GGAGCGTGGCGGCTGCTGTGTGCTCTTCGCGCTTCAACGCCAGGGTATGCGGACTCAGCTGC	515
OY	1096	GAGGAGATGGGCTTTCCTCAGAGGCACTGACCCCATCTCCAGCTTGACGTGCGAAGCGGCGGC	1155
Db	516	GAGGAGATGGGCTTTCCTCAGAGGCACTGACCCCATCTCCAGCTTGACGTGCGAAGCGGCGGC	575
OY	1156	GCCATATGGCATCGTCCGGGCTTCTTCTGTGTGACAGAGGGAGGGCTGCCCAACCCAGAG	1215
Db	576	GCCATATGGCATCGTCCGGGCTTCTTCTGTGTGACAGAGGGAGGGCTGCCCAACCCAGAG	635
OY	1216	CTGCTGAGAGGATCATCTCCGTGTGTATTTGGCCCCAGAGGCGCTTTCTTGAGCGGCATCTGC	1275
Db	636	CTGCTGAGAGGATCATCTCCGTGTGTATTTGGCCCCAGAGGCGCTTTCTTGAGCGGCATCTGC	695
OY	1276	CAGAAGCTGTGGCCGAGGAAAGCTGCCCCGTGACCGCATCTGTGGAGGCGGGACACAGC	1335
Db	696	CAGAAGCTGTGGCCGAGGAAAGCTGCCCCGTGACCGCATCTGTGGAGGCGGGACACAGC	755
OY	1336	TTGGGACCGGTGGCCCGTGGCAAGTCAAGCTTGCGCTATGATGAGACACACTCTGTGTGGGGA	1395
Db	756	TTGGGACCGGTGGCCCGTGGCAAGTCAAGCTTGCGCTATGATGAGACACACTCTGTGTGGGGA	815
OY	1396	TCCCGCTCTCCGCGGGGACCTGGGTGCTGACAGCCGCGCCACTGCTTCCGAGAGGAAACGG	1455
Db	816	TCCCTGCTCTCCGCGGGGACCTGGGTGCTGACAGCCGCGCCACTGCTTCCGAGAGGAAACGG	875
OY	1456	GTCTGTGCTCCGATATGGACGAGTGTGTTGTCGCGGTGCGGCTCAGAGGCTCTCCCAAGCTGTG	1515
Db	876	GTCTGTGCTCCGATATGGACGAGTGTGTTGTCGCGGTGCGGCTCAGAGGCTCTCCCAAGCTGTG	935
OY	1516	CAGCTGGGGGGGTGCAGGCTGTGTCTTACCAAGGGGGCTATCTTCCCTTTCCGGACCCCAAC	1575
Db	936	CAGCTGGGGGGGTGCAGGCTGTGTCTTACCAAGGGGGCTATCTTCCCTTTCCGGACCCCAAC	995
OY	1576	AGCGAGGAGAACAGCAAGATATTTGGCCCGTGCACACTCCAGTCCCTGCGCCCTCA	1635
Db	996	AGCGAGGAGAACAGCAAGATATTTGGCCCGTGCACACTCCAGTCCCTGCGCCCTCA	1055
OY	1636	GAATATCATATCCAGCTGTGTGCTCTCCAGCTGCAGGACCTTGATGATGCAAGATC	1695
Db	1056	GAATATCATATCCAGCTGTGTGCTCTCCAGCTGCAGGACCTTGATGATGCAAGATC	1115
OY	1696	TGTATCGTGAAGGCGCTGGGGGCAACGCAAGTACTATGAGCAACAGGCGGGGTACTCAG	1755
Db	1116	TGTATCGTGAAGGCGCTGGGGGCAACGCAAGTACTATGAGCAACAGGCGGGGTACTCAG	1175
OY	1756	GAGGCTGAGTCCCATATATACGCAATATGTCTGCAATATGCGCTGACTTTCTATGGAAC	1815
Db	1176	GAGGCTGAGTCCCATATATACGCAATATGTCTGCAATATGCGCTGACTTTCTATGGAAC	1235
OY	1816	CAGATCAAGCCCAAGATGTTCTGTACTGTGGCTAACCCGAGAGGTGCAATGATGCTGCAG	1875
Db	1236	CAGATCAAGCCCAAGATGTTCTGTACTGTGGCTAACCCGAGAGGTGCAATGATGCTGCAG	1295
OY	1876	GCGCAGACGGGTGTCTCCCTTTGTGTGAGGACAGCAATCTCTCGAGCGCACGTTGGCGG	1935
Db	1296	GCGCAGACGGGTGTCTCCCTTTGTGTGAGGACAGCAATCTCTCGAGCGCACGTTGGCGG	1355
OY	1936	CTGTGTGGCATTTGTGATGTTGGGGACCTGGCTGTGCTCCCTGCGCCAGAACCGAGGCTCTAC	1995
Db	1356	CTGTGTGGCATTTGTGATGTTGGGGACCTGGCTGTGCTCCCTGCGCCAGAACCGAGGCTCTAC	1415
OY	1996	ACCAATATCAGTGAATTCGCGGAGAGTGAATCTTCAGGCGCATTAAGACATCATCTCCGAAGCC	2055
Db	1416	ACCAATATCAGTGAATTCGCGGAGAGTGAATCTTCAGGCGCATTAAGACATCATCTCCGAAGCC	1475
OY	2056	AGCGGCAATGTGACCCCAAGCTGACCGGATGCTTCTGCTGTGACCGCTCCAGAGGCGCGA	2115
Db	1476	AGCGGCAATGTGACCCCAAGCTGACCGGATGCTTCTGCTGTGACCGCTCCAGAGGCGCGA	1535
OY	2116	GGTGATCCCGGTGTGTGGGATTCACGCTGTGGGCCAGAGATGGACGTTTTTCTTCTTGGGCC	2175
Db	1536	GGTATATCCCGGTGTGTGGGATTCACGCTGTGGGCCAGAGATGGACGTTTTTCTTCTTGGGCC	1595

QY	2176	CGGTTCACAGAGTCCAAAGACACACCTTCCCTCCAGGGTCTCTCTTCCACACAGTGGCGGGCCC	2235
Db	1596	CGGTTCACAGAGTCCAAAGACACACCTTCCCTCCAGGGTCTCTCTTCCACACAGTGGCGGGCCC	1655
QY	2236	ACTGAGCCCCGAGACACACCAACTCAACCTCTCGACCCCATGTAAATATTGTTCGCT	2295
Db	1656	ACTGAGCCCCGAGACACACCAACTCAACCTCTCGACCCCATGTAAATATTGTTCGCT	1715
QY	2296	GTCTGGGACCTCTGTCTAGGTGCCCCCTGATGATGGATGCTCTTTAAATAATAAAGATGG	2355
Db	1716	GTCTGGGACCTCTGTCTAGGTGCCCCCTGATGATGGATGCTCTTTAAATAATAAAGATGG	1775
QY	2356	TTTGTGATT 2363	
Db	1776	TTTTGATT 1783	
RESULT 10			
LOCUS	BC025716		
DEFINITION	Homo sapiens hepsin (transmembrane protease, serine 1), transcript variant 2, mRNA (cDNA clone MGC:34588 IMAGE:5228525), complete cds.		
ACCESSION	BC025716		
VERSION	BC025716.1	GI:19343934	
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, K.H., Shennan, C.M., Schluer, G.D., Altschul, S.F., Zeeberg, B., Bietow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scherzer, T.E., Brownstein, M.J., Ushin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loggellano, N.A., Peters, G.J., Abramson, R.D., Mitaliy, S.J., Bosak, S.A., McMan, P.J., McEran, K.D., Malek, J.A., Gunaratne, P.H., Richards, S., Woley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Vallation, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Faley, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whilton, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butcherfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	PROC. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
PMID	12477932		
REFERENCE	2 (bases 1 to 1761)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: gcgaps-remail.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@nigri.nih.gov Ahter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,		

Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maekeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantiridop, S., Thomas, P.J., Touchman, J.W., Teurigon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 49 Row: 0 Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504480.

FEATURES

Source

Location/Qualifiers

1..1761

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:34588 IMAGE:5228525"

/tissue_type="Pancreas, Spleen, adult pooled"

/clone_1db="NIH_MGC_120"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..1761

/gene="HPN"

/note="Synonym: TMPSR1"

/db_xref="LocusID:3249"

/db_xref="MIM:142440"

198..1451

/gene="HPN"

/codon_start=1

/product="Hepsin (transmembrane protease, serine 1)"

/protein_id="AAH25716.1"

/db_xref="GI:19343935"

/db_xref="LocusID:3249"

/db_xref="MIM:142440"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

/transcript="MAKREGRTVPCSRPVATLTAIGASWAVL"

ORIGIN

Query Match 67.4% Score 1592; DB 9; Length 1761;
Best Local Similarity 99.0% Pred. No. 4,9e-295;

Matches 1612; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

736 CCCCAGGCTCTGCTCCCGTCCATCTCTCAAGGTCCACCTGCGCCAGAGGTCA 795
109 CCTGAGGCTCGCGCCCACTGCTGACCCCA-GGGTCCCACTGCGCCAGAGGTCA 167
796 GCCAGGGAATCATTAACAAGAGCAGTACATGCGCGGAAGAGGGTGGCGACTGTG 855
168 GCCAGGGAATCATTAACAAGAGCAGTACATGCGCGGAAGAGGGTGGCGACTGTG 222
856 CCATGCTCTCAGACCCAGAGTGGCAGCTCTCACTGCGGGGACCTGCTCTTCTGACA 915
228 CCACTGCTCTCAGACCCAGAGTGGCAGCTCTCACTGCGGGGACCTGCTCTTCTGACA 287
916 GCCATCGGGGCGGATCTGCGGCAATTTGGCTGTTCTCTCAGAGTGAACAGAGCGG 975
288 GCCATCGGGGCGGATCTGCGGCAATTTGGCTGTTCTCTCAGAGTGAACAGAGCGG 347
976 CTGATCCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1035
348 CTGATCCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 407
1036 GGGAGTGGCGGCTGCTGCTCTCTGCTCTCAAGCGCCAGAGGTAGCCGACTGACTGC 1095
408 GGGAGTGGCGGCTGCTGCTCTCTGCTCTCAAGCGCCAGAGGTAGCCGACTGACTGC 467

1096 GAGAGATGGGCTTCTTCAAGGCACTGACCCACTCCAGCTGAGAGTGGAGACGGCGGCG 1155
468 GAGAGATGGGCTTCTTCAAGGCACTGACCCACTCCAGCTGAGAGTGGAGACGGCGGCG 527
1156 GCCAATGAGCAGTGGGCTTCTTCTGATGAGCAGAGGAGAGTGGCGCCACACCCAGAGG 1215
528 GCCAATGAGCAGTGGGCTTCTTCTGATGAGCAGAGGAGAGTGGCGCCACACCCAGAGG 587
1216 CTGCTGAGAGTGAATCTCGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1275
588 CTGCTGAGAGTGAATCTCGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 647
1276 CAAATCTGAGGAGGAGAGAGTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1335
648 CAAATCTGAGGAGGAGAGAGTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 707
1336 TTGGAGCGGAGTGGAGAGTGGAGAGTGGAGAGTGGAGAGTGGAGAGTGGAGAGTGGAG 1395
708 TTGGAGCGGAGTGGAGAGTGGAGAGTGGAGAGTGGAGAGTGGAGAGTGGAGAGTGGAG 767
1396 TCCCTGCTCTCGGAGAGCTGGAGTCTGACAGCGCCCACTGCTTCCGAGAGCGAACCAG 1455
768 TCCCTGCTCTCGGAGAGCTGGAGTCTGACAGCGCCCACTGCTTCCGAGAGCGAACCAG 827
1456 GTCTGTGCTCGGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 1515
828 GTCTGTGCTCGGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 887
1516 CAGCTGGAGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 1575
888 CAGCTGGAGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 947
1576 AGCAGAGAGAGAGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1635
948 AGCAGAGAGAGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1007
1636 GAATACATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1695
1008 GAATACATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1067
1696 TGTACCGTGAAGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1755
1068 TGTACCGTGAAGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1127
1756 GAGCTCGAGTCCCATATATGAGATGATGATGATGATGATGATGATGATGATGATGAT 1815
1128 GAGCTCGAGTCCCATATATGAGATGATGATGATGATGATGATGATGATGATGATGAT 1187
1816 CAGATCAAGCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1875
1188 CAGATCAAGCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1247
1876 GGGCAGAGGAGTGGTCCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1935
1248 GGGCAGAGGAGTGGTCCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1307
1936 CTGTGTGAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1995
1308 CTGTGTGAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1367
1996 ACCAAGTCAAGTCACTTCCGAGAGTGAATTTTCAAGGCAATTAAGTCACTCCGAGACC 2055
1368 ACCAAGTCAAGTCACTTCCGAGAGTGAATTTTCAAGGCAATTAAGTCACTCCGAGACC 1427
2056 AGCGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2115
1428 AGCGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1487
2116 GGTGATCCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2175
1488 GGTGATCCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1547
2176 CGGTCCACAGGTCCACAGACCTTCCCTCCAGAGGTCTCTCTTCCACAGTGGCGGCGC 2235

Db	1548	CGGTCACAGGTC	CAAGAGACACCTCCTCCAGGGTCCTCTCTTCCACAGTGGCGGGCCC	1607	
QY	2236	ACTCAGCCCCGAGAC	CAACCAACCTCACCCTCTTGAACCCCAATGTAATATTTGTTCT	2239	
Db	1608	ACTCAGCCCCGAGAC	CAACCAACCTCACCCTCTTGAACCCCAATGTAATATTTGTTCT	1667	
QY	2236	GTCGCGGAC	CTCCTGTCTAGTGTCGCCCTGTATGATGAGGAGTCCTTAAATATTAATAAGATGG	2355	
Db	1668	GTCGCGGAC	CTCCTGTCTAGTGTCGCCCTGTATGATGAGGAGTCCTTAAATATTAATAAGATGG	1727	
QY	2356	TTTTGATT	2363		
Db	1728	TTTTGATT	1735		
RESULT 11					
LOCUS	CQ771486	1769 bp	DNA	linear	
DEFINITION	Sequence 1 from Patent WO2004009803.				
ACCESSION	CQ771486				
VERSION	CQ771486.1	GI:45125533			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1				
TITLE	Xiao, Y.				
JOURNAL	Regulation of human hepsin				
	Patent: WO 2004009803-A 1 29-JAN-2004;				
	Bayer Healthcare AG (DE)				
FEATURES	Location/Qualifiers				
source	1..1769				
	/organism="Homo sapiens"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:9606"				
ORIGIN					
	Query Match	59.5%;	Score 1405;	DB 6;	
	Best Local Similarity	90.0%;	Pred. No. 3.5e-259;	Length 1769;	
	Matches 1592;	Conservative 0;	Mismatches 0;	Indels 177;	
			Gaps 1;		
QY	772	GTCCACCCCTTGSCCCAGAGGTCAGCCAGGGATCATTTAA	CAAGAGGCAATGACATGGCG	831	
Db	1	GTCCACCCCTTGSCCCAGAGGTCAGCCAGGGATCATTTAA	CAAGAGGCAATGACATGGCG	60	
QY	832	CAGAAAGAGGGGTGGCCGGACCTGTGCGATGCTGCTTC	CAGACCAAGAGTGGACCTTCACT	891	
Db	61	CAGAAAGAGGGGTGGCCGGACCTGTGCGATGCTGCTTC	CAGACCAAGAGTGGACCTTCACT	120	
QY	892	GCGGGGACCTGTCTACTTCTTGACAGCCATCGGGGCGGAT	CTCGGACATTTGTGGCTGTT	951	
Db	121	GCGGGGACCTGTCTACTTCTTGACAGCCATCGGGGCGGAT	CTCGGACATTTGTGGCTGTT	180	
QY	952	CTCCTCAGGAATGACACAGAGCCGCTGTAACCAATGACAG	TCATGTCGGAGAGCGCTCGG	1011	
Db	181	CTCCTCAGGAATGACACAGAGCCGCTGTAACCAATGACAG	TCATGTCGGAGAGCGCTCGG	240	
QY	1012	CTCATGTGCTTTTGA	CAAGA	CGGAAGGAGCGTGGCGGCTGTGTGCTCTCTCGCGCTCCAC	1071
Db	241	CTCATGTGCTTTTGA	CAAGA	CGGAAGGAGCGTGGCGGCTGTGTGCTCTCTCGCGCTCCAC	300
QY	1072	GCCAGGGATAGCCGGA	CTCAAGTGTGAGAGATAGGGCTTCTC	AGAGCACTGACCACTCC	1131
Db	301	GCCAGGGATAGCCGGA	CTCAAGTGTGAGAGATAGGGCTTCTC	AGAGCACTGACCACTCC	360
QY	1132	GAGCTGGA	CGTGC	GAACCGCGCGCCCAATGGAACGTCGGGCTTCTTGTGTGGA	CGAG 1191
Db	361	GAGCTGGA	CGTGC	GAACCGCGCGCCCAATGGAACGTCGGGCTTCTTGTGTGGA	CGAG 420
QY	1192	GGAAGGCTGCCCA	CAACCAAGGCTGTGAGGTCACTCCG	TGTGATTTGCCCA	GAG 1251

Db	421	GGGAGGCTGCCCAACACCAGAGGCTGTGAGGTATCTCCGTGTGTATGATGCCCAGA	480
Qy	1252	GGCGTTTCTTGGCCGGCATCTGCA-----	1277
Db	481	GGCGTTTCTTGGCCGGCATCTGCAAGTGAATCTTAAATCAGAACCTCTCTTT	540
Qy	1278	-----	1277
Db	541	AGGCGTTGGGAGGCCAGTCCCTCCAACTCCCAAGATGGGGCCATGTACTTTGAA	600
Qy	1278	-----	1277
Db	601	CCCCCTTAGGGCAGGGCCAAAGCTTGAGCTTGAGGAACTTGAGGCTCAAGTCCCTGTGCGG	660
Qy	1278	-----AGACTGTGGCCGAGGAAGTGGCCCGTGGACCGGATC	1314
Db	661	CCCCCTGCTAACCCTTGTCCCAAGACTGTGGCCGAGGAAGCTGGCCCGTGGACCGGATC	720
Qy	1315	GTGGAGGCGCGGAGCACACAGACTTTGAGCGGATGGCCGTGGCAATCAAGCTTGCCTATGAT	1374
Db	721	GTGGAGGCGCGGAGCACACACTTTGAGCGGATGGCCGTGGCAATCAAGCTTGCCTATGAT	780
Qy	1375	GGAGCANACTGTGTGGGGGATCCCTGCTCTCCGGGGGACTGGGTGCTGACAGCGGCCAC	1434
Db	781	GGAGCANACTGTGTGGGGGATCCCTGCTCTCCGGGGGACTGGGTGCTGACAGCGGCCAC	840
Qy	1435	TGCTTCCCGAGCGGAACCGGATCTGTCCGATGGCGAGTGTGGCCGGTGGCCGCTGGCC	1494
Db	841	TGCTTCCCGAGCGGAACCGGATCTGTCCGATGGCGAGTGTGGCCGGTGGCCGCTGGCC	900
Qy	1495	CAGGCTCTTCCCAACGGTCTGACAGCTGGGGGTGACAGCTTGTGTTACACCGGGGGCTAT	1554
Db	901	CAGGCTCTTCCCAACGGTCTGACAGCTGGGGGTGACAGCTTGTGTTACACCGGGGGCTAT	960
Qy	1555	CTTCCCTTTTGGGAGCCCAACAGCAGAGGAAGAACAGCAACGATTTAGCCCTGTCCACTTC	1614
Db	961	CTTCCCTTTTGGGAGCCCAACAGCAGAGGAAGAACAGCAACGATTTAGCCCTGTCCACTTC	1020
Qy	1615	TCCAGTCCCTGTGCCCTCACAAGATAATCACTCAAGCTGTGTGCTTCCCAAGTGCAGGCTCAG	1674
Db	1021	TCCAGTCCCTGTGCCCTCACAAGATAATCACTCAAGCTGTGTGCTTCCCAAGTGCAGGCTCAG	1080
Qy	1675	GGCCTGTGTGATGGCAAGATCTTAAOCGTGAAGGGGCTGGGGGCAACACGAGTACTTATGGC	1734
Db	1081	GGCCTGTGTGATGGCAAGATCTTAAOCGTGAAGGGGCTGGGGGCAACACGAGTACTTATGGC	1140
Qy	1735	CAACAGGCTCGGGGTACTCCAGAGGAGCTCGATCCCAATATCAGCAATGATGTCTGCAAT	1794
Db	1141	CAACAGGCTCGGGGTACTCCAGAGGAGCTCGATCCCAATATCAGCAATGATGTCTGCAAT	1200
Qy	1795	GGCGCTGACTTTTATGGAACCAAGATCAAGCCCAAGATTTCTGTGTGCTTACCCCGAG	1854
Db	1201	GGCGCTGACTTTTATGGAACCAAGATCAAGCCCAAGATTTCTGTGTGCTTACCCCGAG	1260
Qy	1855	GGTGGCAATTGATGCTGSCCAGGGGCGACAGGGGTGTGCTCCCTTTGTGTGAGCAGAGATC	1914
Db	1261	GGTGGCAATTGATGCTGSCCAGGGGCGACAGGGGTGTGCTCCCTTTGTGTGAGCAGAGATC	1320
Qy	1915	TCTTGGACGCGCACGTTTGGCGGCTGTGTGTGAGATTTGGGACACTGTGCTGTGCCCTG	1974
Db	1321	TCTTGGACGCGCACGTTTGGCGGCTGTGTGTGAGATTTGGGACACTGTGCTGTGCCCTG	1380
Qy	1975	GGCCAGAAAGCAGGCGCTTACCAAAAGTCAAGTCACTTCCGGGAGTGGATCTTCCAGGGC	2034
Db	1381	GGCCAGAAAGCAGGCGCTTACCAAAAGTCAAGTCACTTCCGGGAGTGGATCTTCCAGGGC	1440
Qy	2035	ATTAAGAATCACTCCGAAGCAGCGGATGTGTACCCAGCTTGAACCGAGTGTCTTCGC	2094
Db	1441	ATTAAGAATCACTCCGAAGCAGCGGATGTGTACCCAGCTTGAACCGAGTGTCTTCGC	1500
Qy	2095	TGGCAGCTTCCAGGGCCGAGGTGATCCCGGTGTGTGGATTCACAGTGTGGCCGAGGATG	2154
Db	1501	TGGCAGCTTCCAGGGCCGAGGTGATCCCGGTGTGTGGATTCACAGTGTGGCCGAGGATG	1560

QY 2155 GGACGTTTCTTCTTGGGCCCCGCTCAAGGTCACAGTCCAAAGACACCTCCCTCCAGAGTCT 2214
DB 1561 GGAGCTTTTCTTCTTGGGCCCCGCTCAAGTCCAAAGACACCTCCCTCCAGAGTCT 1620
QY 2215 CTCTTCCACAGTGGGCGGCCCCACTCAGCCCCGAGACCAACCACTCACTCCCTGAGACC 2274
DB 1621 CTCTTCCACAGTGGGCGGCCCCACTCAGCCCCGAGACCAACCACTCACTCCCTGAGACC 1680
QY 2275 CCATGTAATAATTTGTTCTGCTGTCTGGGACTCTGTCTAAGTGTCCCTGATGATGGATG 2334
DB 1681 CCATGTAATAATTTGTTCTGCTGTCTGGGACTCTGTCTAAGTGTCCCTGATGATGGATG 1740
QY 2335 CTCTTAAATAATAAGATGTTTGATT 2363
DB 1741 CTCTTAAATAATAAGATGTTTGATT 1769

RESULT 12
AR255885 1615 bp DNA linear PAT 20-DEC-2002
LOCUS AR255885 Sequence 1 from patent US 6482630.
DEFINITION AR255885
ACCESSION AR255885
VERSION AR255885.1 GI:27305128
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1615)
AUTHORS Gan, W., Ye, J., DiFrancesco, V. and Beasley, E.M.
TITLE Isolated human protease proteins, and uses thereof
JOURNAL human protease proteins, and uses thereof
FEATURES Patent: US 6482630-A 1 19-NOV-2002;
source Location/Qualifiers
1..1615
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 53.7%; Score 1269; DB 6; Length 1615;
Best Local Similarity 91.1%; Pred. No. 4.2e-233;
Matches 1411; Conservative 0; Mismatches 15; Indels 123; Gaps 1;

QY 815 GAGCAGTGAATGCGCCGAGAAAGAGGTGGCCGAGACTGTGCTGCTTCAGACCA 874
DB 167 GTGGCCCCAGATGTGTCTGTTCAGAGTGGCCGAGACTGTGCTTCAGACCA 226
QY 875 AGGTGGCAGCTCTCACTGCGGGGACCCGCTACTTCTGACAGCAATCGGGGCGGCACTCT 934
DB 227 AGGTGGCAGCTCTCACTGCGGGGACCCGCTACTTCTGACAGCAATCGGGGCGGCACTCT 286
QY 935 GGGCATTGTGGCTGTCTCTCTCAGAGTGAACAGAGCCGCTGTACCCAGTGCAGTCA 994
DB 287 GGGCATTGTGGCTGTCTCTCTCAGAGTGAACAGAGCCGCTGTACCCAGTGCAGTCA 346
QY 995 GCTCTGCGGAGCTCTGCTCATGTCTTTGACAAAGCCGAGACGCTGGGCTGCTGT 1054
DB 347 GCTCTGCGGAGCTCTGCTCATGTCTTTGACAAAGCCGAGACGCTGGGCTGCTGT 406
QY 1055 GCTCTGCGGCTCCAAAGCCGAGGTAAGCAGTCACTGAGTGCAGAGATGGGCTTCTCA 1114
DB 407 GCTCTGCGGCTCCAAAGCCGAGGTAAGCAGTCACTGAGTGCAGAGATGGGCTTCTCA 466
QY 1115 GGGCAGTGAACCACTCCAGAGTGAAGTGCAGACGCGGGGCCAATGGCAGTCCGCT 1174
DB 467 ----- 466
QY 1175 TCTTCTGTGTGAGACGAGGAGAGCTGCCACACCCAGAGGCTGTGAGAGTCACTCCG 1234
DB 467 ----- 466
QY 1235 TGTGTATTTGCCCCAGAGCGGTTTCTTGGCGGCAATCTGCACAAGACTGTGGCGCAGGA 1294

DB 467 ---GATATGCCCCAGAGGCGGTTTCTTGGCGGCAATCTGCACAAGACTGTGGCGCAGGA 523
QY 1295 AGCTGCCGTGTGACCGCATCTGTGGAGGCGCGGACACCACTTGGGCGGTGGCGGTGGC 1354
DB 524 AGCTGCCGTGTGACCGCATCTGTGGAGGCGCGGACACCACTTGGGCGGTGGCGGTGGC 583
QY 1355 AAGTCAAGCTTGGTATGATGAGAGACACCTGTGGGGGAGATCCGTGCTCCGGAGACT 1414
DB 584 AAGTCAAGCTTGGTATGATGAGAGACACCTGTGGGGGAGATCCGTGCTCCGGAGACT 643
QY 1415 GGGTGTGACAGCCGCCCACTGTTCCCGAGAGCGGAACCGGGTCTGTCCGATGGCAG 1474
DB 644 GGGTGTGACAGCCGCCCACTGTTCCCGAGAGCGGAACCGGGTCTGTCCGATGGCAG 703
QY 1475 TGTGTGCGGTGTGGTGGCCCAAGGCTCTCCCAAGGTTCTCAGTGTGGGGGTGACAGCTG 1534
DB 704 TGTGTGCGGTGTGGTGGCCCAAGGCTCTCCCAAGGTTCTCAGTGTGGGGGTGACAGCTG 763
QY 1535 TGTGTCAACAGAGGGGGGATCTTCCCTTGGGAGCCCAACAGGAGAGGAACAGCAAG 1594
DB 764 TGTGTCAACAGAGGGGGGATCTTCCCTTGGGAGCCCAACAGGAGAGGAACAGCAAG 823
QY 1595 ATATTGCGCTGTGTCACTCTCAGTCCCTGCGCTTCAAGAAATACATCCAGCTGTGT 1654
DB 824 ATATTGCGCTGTGTCACTCTCAGTCCCTGCGCTTCAAGAAATACATCCAGCTGTGT 883
QY 1655 GCTTCCAGCTGCGGCGCAGGCGCTGTGTGATGAGCAAGATTTGATCCGTGACGGGCTGGG 1714
DB 884 GCTTCCAGCTGCGGCGCAGGCGCTGTGTGATGAGCAAGATTTGATCCGTGACGGGCTGGG 943
QY 1715 GCAACAGCAGTACTATGAGCCCAAGGCGGGGATCTCCAGAGAGCTGATCCCAATPA 1774
DB 944 GCAACAGCAGTACTATGAGCCCAAGGCGGGGATCTCCAGAGAGCTGATCCCAATPA 1003
QY 1775 TCAGCAATGATGTCTGCAATGCGCTGACTTCTATGAAACCAATCAAGCCCAAGATGT 1834
DB 1004 TCAGCAATGATGTCTGCAATGCGCTGACTTCTATGAAACCAATCAAGCCCAAGATGT 1063
QY 1835 TCTGTCTGTGCTACCCCGAGGGGTGCAATTGATGCTTGCAGGGGACAGCGGTGCTCT 1894
DB 1064 TCTGTCTGTGCTACCCCGAGGGGTGCAATTGATGCTTGCAGGGGACAGCGGTGCTCT 1123
QY 1895 TTTGTGTGAGGAGAGATCTCTGAGCGCCACGATGCGGCTGTGTGGCAATTGTGACTT 1954
DB 1124 TTTGTGTGAGGAGAGATCTCTGAGCGCCACGATGCGGCTGTGTGGCAATTGTGACTT 1183
QY 1955 GGGCAGCTGTGCTGCTGCGCCAGAGAGCCAGAGGCTCTACACCAAGTCAAGTCACTTCC 2014
DB 1184 GGGCAGCTGTGCTGCTGCGCCAGAGAGCCAGAGGCTCTACACCAAGTCAAGTCACTTCC 1243
QY 2015 GGGAGTGAATCTTCAAGGCCATTAAGAATCACTCCGAAGCCAGCGGCAATGTGACCCAGC 2074
DB 1244 GGGAGTGAATCTTCAAGGCCATTAAGAATCACTCCGAAGCCAGCGGCAATGTGACCCAGC 1303
QY 2075 TCTGACGGTGTGCTTCTGCGGCGACGCTCCAGAGGCGGAGAGTATCCCGTGTGGGA 2134
DB 1304 TCTGACGGTGTGCTTCTGCGGCGACGCTCCAGAGGCGGAGAGTATCCCGTGTGGGA 1363
QY 2135 TCCAGCGCTGGGCGAGAGTGGAGCTTTTCTTGTGGGCCCGGTTCACAGAGTCCAAAGA 2194
DB 1364 TCCAGCGCTGGGCGAGAGTGGAGCTTTTCTTGTGGGCCCGGTTCACAGAGTCCAAAGA 1423
QY 2195 CACCTTCTCTCAGAGGCTCTCTTCCACAGTGGCGGAGCCCACTCAGCGCCGAGACCAAC 2254
DB 1424 CACCTTCTCTCAGAGGCTCTCTTCCACAGTGGCGGAGCCCACTCAGCGCCGAGACCAAC 1483
QY 2255 CAACCTCAACCTCTGACCCCAAGTAAATATTTGTGTGCGTGTGGAGACTCGTGTAG 2314
DB 1484 CAACCTCAACCTCTGACCCCAAGTAAATATTTGTGTGCGTGTGGAGACTCGTGTAG 1543
QY 2315 GTGCCCTGTATGATGAGATGCTCTTTAAATATTAAGATGTTTGATT 2363
DB 1544 GTGCCCTGTATGATGAGATGCTCTTTAAATATTAAGATGTTTGATT 1592

RESULT 13
LOCUS HSHPEPL 1199 bp mRNA linear PRI 22-SEP-1995
DEFINITION H.sapiens liver mRNA for serine protease hepsin.
ACCESSION X07002
KEYWORDS X07002.1 GI:32065
SOURCE hepsin; membrane protein; serine protease; zymogen.
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 1199)
AUTHORS Leytus,S.P., Loeb,K.R., Hagen,F.S., Kurachi,K. and Davie,E.W.
TITLE A novel trypsin-like serine protease (hepsin) with a putative
JOURNAL transmembrane domain expressed by human liver and hepatoma cells
MEDLINE Biochemistry 27 (3), 1067-1074 (1988)
PUBMED 88209431
COMMENT 2835076
FEATURES
source see X07732 for hepatoma hepsin full length cDNA.
Location/Qualifiers
1..1199
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HUM 1250"
/issue_type="liver"
/clone_id="pBR322"
1..1199
/codon_start=1
/product="hepsin"
/protein_id="CAA30058.1"
/db_xref="GI:899286"
/db_xref="UniProt/Swiss-Prot:P05981"
/translation="MSGPFCVDEGRILPHOTRLLEVLIVCCPRRFLAATICODGRK
LPVDRIVIGARDTSLGKPMQVSLRDGHLGSLSDWVLTAAHFPERRRLSLW
RFAVAQAQSPHQLQSVQAVVTHGGLPRDPSSENSIDALVLSPLPTEYI
QPVCLPAAGALVDGKI CTYMGNTQYGGQAVLQEARVPIISNVCAGADYGAQ
IKPMFCAGYBEGGIDACQDSGSPFYCEDSISRTPRWRILGIVSMGTGALAKPKGV
YTKVSDPREWIFQAIKTHSEASGAVTOL"
sig_peptide <1..147
mat_peptide 148..902
polya_site 1199
ORIGIN
Query Match 50.7%; Score 1199; DB 9; Length 1199;
Best local similarity 100.0%; Pred. No. 1.2e-219; Indels 0; Gaps 0;
Matches 1199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1165 ACGTCGAGCTTCTTCTGTGTGAGCAGAGGAGGCTGCCACACCCAGAGGCTCTGAG 1224
Db 1 ACGTCGAGCTTCTTCTGTGTGAGCAGAGGAGGCTGCCACACCCAGAGGCTCTGAG 60
QY 1225 GTATCTCCGCTGTGTATGTCGCCAGAGGCGCTTTCTTGGCGGCACTCTGCAAGACTGT 1284
Db 61 GTATCTCCGCTGTGTATGTCGCCAGAGGCGCTTTCTTGGCGGCACTCTGCAAGACTGT 120
QY 1285 GAGCGCAGAGAGCTGCGCGTGGACCGCATCTGAGGAGGCGGAGACACAGCTTGGCGCG 1344
Db 121 GAGCGCAGAGAGCTGCGCGTGGACCGCATCTGAGGAGGCGGAGACACAGCTTGGCGCG 180
QY 1345 TGCGCGTGGCAAGTCAAGCTTGGTATGATGAGACACACTCTGTGAGGAGATCCCTGCTC 1404
Db 181 TGCGCGTGGCAAGTCAAGCTTGGTATGATGAGACACACTCTGTGAGGAGATCCCTGCTC 240
QY 1405 TCCGGGAGACTGGGCTGTGACGCGCCCACTGCTTCCCGAGCGGAAACCGGGTCTCTGTC 1464
Db 241 TCCGGGAGACTGGGCTGTGACGCGCCCACTGCTTCCCGAGCGGAAACCGGGTCTCTGTC 300
QY 1465 CGATGCGAGAGTGTTCGCGGTGCGGTGGCCAGGCTCTCTCCCAAGGCTCTGCAAGTGGG 1524

Db 301 CGATGCGAGAGTGTTCGCGGTGCGGTGGCCAGGCTCTCTCCCAAGGCTCTGCAAGTGGG 360
QY 1525 GTGCAAGCTGTGTGTCTACACAGGGGCTATCTTCCCTTTGGAGACCCCAACAGGAGAG 1584
Db 361 GTGCAAGCTGTGTGTCTACACAGGGGCTATCTTCCCTTTGGAGACCCCAACAGGAGAG 420
QY 1585 AACGACAGATATATGCGCTGTGTACCTCTCCAGTCCCTGCGCCCTCAGAGATATATC 1644
Db 421 AACGACAGATATATGCGCTGTGTACCTCTCCAGTCCCTGCGCCCTCAGAGATATATC 480
QY 1645 CAGCTGTGTGTCTCCAGTCCCTGCGCCCTCAGAGGCTGTGTATGAGCAAGTCTGATCCGTG 1704
Db 481 CAGCTGTGTGTCTCCAGTCCCTGCGCCCTCAGAGGCTGTGTATGAGCAAGTCTGATCCGTG 540
QY 1705 ACGGCTGT 1764
Db 541 ACGGCTGT 600
QY 1765 GTCCCATATATGAGATATGATGT 1824
Db 601 GTCCCATATATGAGATATGATGT 660
QY 1825 CCAAGATGT 1884
Db 661 CCAAGATGT 720
QY 1885 GT 1944
Db 721 GT 780
QY 1945 ATTGTGATGT 2004
Db 781 ATTGTGATGT 840
QY 2005 AGTGAATTTCCGAGAGT 2064
Db 841 AGTGAATTTCCGAGAGT 900
QY 2065 GTGACCAAGCTGT 2124
Db 901 GTGACCAAGCTGT 960
QY 2125 GGT 2184
Db 961 GGT 1020
QY 2185 GGT 2244
Db 1021 GGT 1080
QY 2245 CGAGACCAAGCTGT 2304
Db 1081 CGAGACCAAGCTGT 1140
QY 2305 TCCGT 2363
Db 1141 TCCGT 1199

RESULT 14
LOCUS RNHEPA 1739 bp mRNA linear ROD 19-JUL-1993
DEFINITION R.norvegicus mRNA for hepsin.
ACCESSION X70900
VERSION X70900.1 GI:57928
KEYWORDS hepsin; protease; serine protease.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
REFERENCE 1
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

AUTHORS Farley, D., Raymond, F. and Nick, H.
 TITLE Cloning and sequence analysis of rat hepsin, a cell surface serine
 JOURNAL Biochim. Acta 1173 (3), 350-352 (1993)
 MEDLINE 93305733
 PUBMED 8318546
 REFERENCE 2 (bases 1 to 1739)
 AUTHORS Farley, D.L.
 TITLE Direct Submission
 JOURNAL Submitted (28-JAN-1993) D.L. Farley, Ciba Geigy, K125.117, 4002
 BASEL, SWITZERLAND

FEATURES
 source
 1..1739
 location/Qualifiers
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague Dawley"
 /db_xref="taxon:10116"
 /sex="male"
 /tissue_type="liver"
 /dev_stage="7-8 week"
 1..1739
 /gene="hepsin"
 1..1739
 /evidence="experimental"
 185..1435
 /gene="hepsin"
 /codon_start=1
 /product="hepsin"
 /protein_id="CAA50256.1"
 /db_xref="GI:57929"
 /db_xref="GOL:005511"
 /db_xref="UniProt/Swiss-Prot:Q05511"
 /translation="MAKGGRTAPCCSRPKVAALTVGLFLTGIGAASMAVITLLR
 SDOELVQVLSPEBSRLVLDKTEGTRILCSRSNRAVGLGSEMGFLALAHSE
 LDVRTAGANSGFCEVDEGLPLAORLDVLSVDCDPRGFLPATCODGRRLPVD
 RIVGGDSILGRMPQVSLRYDGLTHCGSILSGDWLTALCPERRRVLSSRRVPA
 GAVARTSRHAYVQLQAVIYHGYILPRDPITDENSNDIALVHSSILPLEYIOPVC
 LPAAGQALVDKVCVTGWTGNTQFTGQDAVVLQEARVPTISNEVNSDFGNDIKPK
 MFCAGYPEGIDACGDSGGHFCEDRISGTSRMLCGIVSMGTGCAIARPKGVYTVV
 IDPREWIFQAIKTHSEATGMVTP"
 185..667
 /gene="hepsin"
 /gene="hepsin"
 668..1432
 /gene="hepsin"
 /product="hepsin"

ORIGIN
 Query Match 43.3%; Score 1023.6; DB 10; Length 1739;
 Best Local Similarity 79.3%; Pred. No. 4.6e-186;
 Matches 1314; Conservative 0; Mismatches 309; Indels 35; Gaps 7;

QY 732 GAGTCCCCAGAGCCCTGCTCCCGCTCATCTCTCAGAGGTCACACCTGAGCCAGAG 791
 DB 91 GACTAACCCCAACTGACCATCTCCGCGAACCCAGGGTTCGCCCCAGCCCAAG 150
 QY 792 GTCAAGCAGGAATCATTAAAGAGAGTGCATGCGCAGAAAGAGGTGCGCGAG 851
 DB 151 GTCAACCTGGGAATCATTTAAAGAGTCCCTGACATGCGC--GAAAGAGGGTGGCCGAG 207
 QY 852 TGTGCGATGCTGCTCCAAACCCAGAGTGGCAGCTTCACTGCGGGAGACCTGCTACTTCT 911
 DB 208 TGACCAATGCTGCTCCAAACCCAGAGTGGCAGCTTCACTGCGGGAGACCTGCTACTTCT 267
 QY 912 GACAGCATCGGGGCGGAGCTCTGGGCAATGTGAGCTGTTCTCTCAGAGTACAGAGA 971
 DB 268 GACAGCATTTGGGGCTGCGCTCTGGGCAATGTGAGCTTCTCTCAGAGTACAGAGA 327
 QY 972 GCCGCTGTAACCAAGTGCAGAGTCACTGCGGAGCGCTCGGCTCATGTGCTTTGACAAAG 1031
 DB 328 GCCATGTAACCAAGTGCAGAGTCACTGCGGAGCGCTCGGCTCATGTGCTTTGACAAAG 387
 QY 1032 GGAAGGAGACGTGGCGGCTGCTGCTCTCGGCTTCAACCCAGAGGTAGCCGAGCTAG 1091

DB 388 AGAGGAAACGTGAGAGCTGCTGTGCTCTTCAACGCTCCAAACCCAGAGGTAGAGAGGCTCG 447
 QY 1092 CTGCGAGAGATGGGCTTCTCTCAGAGCACTGACCACTCCAGAGTGCAGAGCGC 1151
 DB 448 CTGTGAGAGATGGGCTTCTCTCAGAGCACTGACCACTCCAGAGTGCAGAGCGC 507
 QY 1152 GGGCGCCAAATGGCACGTGGGCTTCTCTGTGTGAGCAGAGGGAGAGGCTGCCACACCA 1211
 DB 508 GGGCGCCAAATGGCACATCGGGCTTCTCTGTGTGAGCAGAGGGAGAGGCTGCCCTCA 567
 QY 1212 GAGGCTGCTGAGAGTCACTCCGTGTGTGATTTGCCAGAGGCGGCTTCTTGGCGCCAT 1271
 DB 568 GCGGTGCTGATGTCACTCTGTATGAGCACTGTCTTAGAGCCGATTTCTGACTGCGAC 627
 QY 1272 CTGCCAAGACTGTGGCCGACAGAAAGCTGCCGTGACCGCATCTGTGGAGGCGGAGAC 1331
 DB 628 CTGCCAAGACTGTGGCCGACAGAAAGCTGCCGTGATGCAATTTGGGGGGCAGAGACG 687
 QY 1332 CAGCTTGGGCGGCTGGCCGCAAGTCAAGCTTTCCTATATGATGAGCACACTTGTGG 1391
 DB 688 CAGCTTGGGAGATGGCCATGCGACAGTCAAGCTTTCCTATATGATGAGCACACTTGTGG 747
 QY 1392 GGGATCCCTGCTCCGGGAGACTGGGTGCTGACAGCGCCCACTGCTCCCGAGCGGA 1451
 DB 748 GGGATCCCTGCTGCTCCGGGAGACTGGGTGCTGACAGCGCCCACTGCTCCCGAGCGGA 807
 QY 1452 CCGGGCTCTGTCCCGATGGCGAGTGTGGCGGTGCGGTGCGCCAGGCTTCCCGACGG 1511
 DB 808 CCGGGCTCTGTCTCGGTGGCGAGTATTTGCTGCTGTAGCTCCGGACCTCAGCTCATGC 867
 QY 1512 TCTGCACTGGGGGTGACAGGCTGTGTGTTACCAAGGGGGCTATCTTCCCTTTGGGAGCC 1571
 DB 868 CGTCAGCTGGGGGTTCAGGCTGTGATCTAATGAGGGGCTACCTTCCCTTTGAGAGCC 927
 QY 1572 CAACAGGAGAGAGAAACAGACAGATGATTTGACCTGCTCCAGCTTCCAGTCCCTGCGCCCT 1631
 DB 928 TACTATGACAGAAACAGACATGATTTGACCTGCTCCAGCTTCCAGTCCCTGCGCCCT 987
 QY 1632 CACAGAAATACATCCAGCTGTGTGCTCCAGCTGCGCGCCAGGCGCTGTGATGAGCA 1691
 DB 988 CACAGAAATACATCCAGCTGTGTGCTCCAGCTGCGCGCCAGGCGCTGTGATGAGCA 1047
 QY 1692 GATCTGTACCTGACCGGCTGTGGGACACAGCAAGTACTATGAGCCACAGGCGGGGTACT 1751
 DB 1048 GGTCTGTACAGTGAACCGGCTGTGGGTAACACAGTTCATATGAGCCAGAGCTGTGTCT 1107
 QY 1752 CCAGAGGCTGAGTCCCATTAATCAGCAATGATGTGCAATGAGCGCTGACTTATAG 1811
 DB 1108 CCAAGAGGCGCGGATCCCATTAATGAGCAATGATGTGCAATGAGCGCTGACTTATAG 1167
 QY 1812 AAACAGATCAAGCCCAAGATGTTCTGTGCTGAGTACCCCGAGGGTGCATTTGATGCTG 1871
 DB 1168 GAATCAATCAAAACCAAGATGTTCTGTGCTGAGTACCCCGAGGGTGCATTTGATGCTG 1227
 QY 1872 CCAGGCGACAGCGGTGCTCTTGTGTGTGAGAGACAGATCTCTGAGACGCAAGTTG 1931
 DB 1228 CCAGGCGACAGCGGTGCTCTTGTGTGTGAGAGACAGATCTCTGAGACGCAAGTTG 1287
 QY 1932 GCGGCTGTGTGAGTGTGAGTTGGGGCACTGGCTGTGCGCTTGGCCAGAGAGCGAGGCT 1991
 DB 1288 GCGGCTGTGTGAGTGTGAGTTGGGGCACTGGCTGTGCGCTTGGCCAGAGAGCGAGGCT 1347
 QY 1992 CTACACCAAGTCAAGTCACTTCCGAGAGTGAATCTTCCAGGCAATTAAGACTCACTCGA 2051
 DB 1348 GTACACCAAGTCAAGTCACTTCCGAGAGTGAATCTTCCAGGCAATTAAGACTCACTCGA 1407
 QY 2052 AGCCAGCGGATGTGATCCAGCTTGAAC-----CGGTGCTTCTGCTGC-6CA 2100
 DB 1408 AGCTACCGGATGTGATCTAGCCCTGACCCCGGCTCATGCTGCTCCGCGCTGCTCA 1467
 QY 2101 GCGTCAAGGGCGGAGGTATC-----CCGGTGGTGGAGTCCAGCGCTGG 2145

Db 1468 GCATCCAGAGTCAGAGTTGCTCTGAGGCTCCAGCCGCACTGAGGCTCCACACTGAG 1527

Qy 2146 CCGAGAGTGGAGCTTTTCTTTCTTTGGCCGGCTCCAGGTCCAGAGCACTCTCCTC 2205

Db 1528 CTTCACTAGAGAGGTTTCTGCTGAGTCCAGTCCAGATCCAGAG--ATGCTGAGTTC 1585

Qy 2206 CAGGATCCTCTCTTCCAGAGTGGGGCCCACTAGCCCGGACCACTCACTCAACC 2265

Db 1586 CAAAGACCTCTCTTCCAGAGTGGCCGCGCACTCAATCCAGGGCCATGGGCTCACCTT 1645

Qy 2266 TCTGACCCCATGTAATATTGTTCTGCTGCTGAGACTCTGTCTAGTGGCCCTGAT 2325

Db 1646 CCC--ACCCCATGTAATATTACTCTGCTGAGG--GGCTGCTTTCAGAGCGCCCTT 1701

Qy 2326 GATGAGATGCTCTTAATAATAAGATGTTGATT 2363

Db 1702 GTGCGAGTCTCTTAATAATAAGGTTTGATT 1739

RESULT 15

AF030065 1781 bp mRNA linear ROD 12-NOV-1997

LOCUS AF030065 Mus musculus serine protease hepsin mRNA, complete cds.

DEFINITION AF030065

VERSION AF030065.1 GI:2606036

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.

AUTHORS 1 (bases 1 to 1781)

TITLE Vu,T.-K.H., Liu,R.W., Haakema,C., Tomasek,J.J. and Howard,E.W. Identification and cloning of the membrane-associated serine protease, hepsin, from mouse preimplantation embryos

JOURNAL J. Biol. Chem. (1997) In press

REFERENCE 2 (bases 1 to 1781)

AUTHORS Vu,T.-K.H., Liu,R.W., Haakema,C., Tomasek,J.J. and Howard,E.W. Direct Submission

JOURNAL Submitted (14-OCT-1997) Pathology, University of Oklahoma Health Sciences Center, 940 Stanton L. Young Blvd, Oklahoma City, OK 73104, USA

FEATURES

source location/Qualifiers

1..1781

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/tissue_type="liver"

205..1455

/note="serine protease"

/codon_start=1

/product="hepsin"

/protein_id="AAB84221.1"

/db_xref="GI:2606037"

/translation="MAKEGRTAACSRPKVALIVGTLFLTGIGASMAIVITILQ SDEPLVOVLSPEDSRLAVLDKTEGTRILCSRSNARVAGLCEBEMFLALSHB LDVTAGANGSPFCVDEGLPLAORLDVITSVCDPRGFLFATCCODGRLLPVD RIVGGDSSISGRMVOVSLRYDTHLGGSLISGDWLTAAHCPRENRRLVSRVVA GAVARTSPHVAVOLGVAIVYHGTLPRPDRTIDNSMDIALVHSSLPITYIQPC LPAAGQALVDKIVCTVTGMENTQFYGOQAMLOEARVPIISNEVCNSPDTGQIKK MFCAGYPEGGIDACQSDGPFVCEDSISGTSRMLGIVSWGALCARPKGVYTRY TDFREWI.FKIKTHSEASGMVTP"

ORIGIN

Query Match 43.3%; Score 1022.2; DB 10; Length 1781;

Best Local Similarity 79.2%; Pred. No. 8.6e-186;

Matches 1314; Conservative 0; Mismatches 308; Indels 37; Gaps 7;

Db 171 GTCAACTGGGAATCATTAACAAGAGTCCCTGACATGGC---GAAAGAGGGTGGCCGAG 227

Qy 852 TGTGCCATGTGCTCCAGACCCAAAGTGGAGCTTCTACGTGGGGACCCCTGACTTCT 911

Db 228 TGCAGCATGTGCTCCAGACCCAAAGTGGAGCTTCTATTGTGGTACCCCTGCTTCT 287

Qy 912 GACAGCATGGGGGCGGATCTGGGGCATTTGGTGTCTTCTCTCAGAGTGAACAGGA 971

Db 288 GACAGGATTTGGGGCCCGGCTCTGGGGCATTTGTACATCTTACTGACAGTGAACAGGA 347

Qy 972 GCCGCTGATCCAGTGCAGGTCAAGCTGCGGACGGTGGGCTCATGTCTTTGACAGAC 1031

Db 348 GCCACTGTACCAAGTGCAGGTCAAGTCCAGGGGACCTCAAGCTTGCAGTGTGACAGAC 407

Qy 1032 GGAAGGAGCGTGGCGGTGTGCTGTCTGCGCTCCAAAGGCGAGGGTACGCGGACTCAG 1091

Db 408 GGAAGGATCGTGGAGGCTACTGTCTCTACGCTCCAAAGTCCAGGGTGGAGGGCTTCGG 467

Qy 1092 CTGCGAGAGATGGGCTTCTCAGAGGCACTGACCCACTCCGAGCTGAGACGTGGACGCG 1151

Db 468 CTGTGAGAGATGGGCTTCTCAGAGGCTCTGGGCACTGGAGGTGGAATGGGCACTGC 527

Qy 1152 GGGGCGCAATGACAGTGGGCTTCTTGTGTGACGAGGAGAGGCTGCCACACCA 1211

Db 528 GGGGCGCAATGACAGTGGGCTTCTTGTGTGACGAGGAGGAGTGGCTGCTCA 587

Qy 1212 GAGGCTGTGAGGTACTCTCGGTGTGATGTCGCCAGAGGCGCTTCTTGGCGGCTAT 1271

Db 588 GAGGCTGTGAGGTACTCTCTGTATGTGACTGCTCAGAGGCGGATTCCTGACTGACAC 647

Qy 1272 CTGCGCAAGATGTGCGCGCAGAGAGCTGCCGTGACCGGATGCTGAGAGGCGCGGACAC 1331

Db 648 CTGCGCAAGATGTGCGCGCAGAGAGCTGCCGTGAGCGGATGTGGGGGCGCAGACAG 707

Qy 1332 CAGTTGGCGCGGTGCGCGTGGCAAGTCAAGCTTCTGATGATGAGACACACTCTGTGG 1391

Db 708 CAGTGTGGGAAGGTGGCGGTGGCAGAGTCAAGCTTCTGATGATGAGAGCCCACTCTGTGG 767

Qy 1392 GGGATCCCTGCTCTCGGGGAGTGGGTGCTGACAGCGGCCCACTGCTTCCCGAGGGGAA 1451

Db 768 GGGATCCCTGCTCTCGGGGAGTGGGTGCTGACAGCTGACATTTGCTTCCAGAGCGGAA 827

Qy 1452 CCGGATCTGCTCCGATGAGGAGTGTGGCGGTGCGGTGCGGACCGGCTCTCCCAAG 1511

Db 828 CCGGATCTGCTCTCGGAGGAGATTTGTGTGATGCTGATCCCGAGCTCACCCATGC 887

Qy 1512 TGTGCACTGGGGTGGAGGCTGTGCTTCAACAGGGGCTATCTTCTTGGGAGCC 1571

Db 888 TGTGCAACTGGGGTTCAGAGCTGTGATCTATATGGGGGCTACCTTCTTGGAGACC 947

Qy 1572 CAACAGCGAGAGAGAACGAAAGATATGCGCTGGTCACTCTCCAGTCCCTGACCCCT 1631

Db 948 TACTATTGAGAAAGACGATGACATGTGCTTGTGCTCACTCTTACGCTCTGCTCT 1007

Qy 1632 CACAGATATATCATCAGCTGTGTGCTTCCAGCTGCGGCGGACCGCTGTGTGACAA 1691

Db 1008 CACAGATATATCATCAGCTGTGTGCTTCCAGCTGCGGCGGACCGCTGTGTGACAA 1067

Qy 1692 GATCTGTACCTGTACGAGGCTGGGACACAGCTACTATGCGCAACAGGCGGGGACT 1751

Db 1068 GGTCTGTACTGTACAGCGGCTGGGGTAAACACAGTTCTATAGGCGCAACAGGCTATGTGCT 1127

Qy 1752 CCAAGAGGCTGAGTCCCATTAATACGATGATGTGTGCAATGAGGCTGACTTCTATGG 1811

Db 1128 CCAAGAGGCTGAGTCCCATTAATACGATGATGTGTGCAATGAGGCTGACTTCTATGG 1187

Qy 1812 AAACAGATCAACCCCAAGATGTTCTGTGTGCTGCTACCCCGAGGCTGGCAATGACTG 1871

Db 1188 GAATCAATCAACCCCAAGATGTTCTGTGTGCTGCTACCCCGAGGCTGGCAATGACTG 1247

Qy 1872 CCAAGGCGACAGCGGTGTCTTTGTGTGTGAGACAGATCTCTGACGCGACGTTG 1931

Db 1248 CCAAGGCGACAGCGGTGTCTTTGTGTGTGAGACAGATCTCTGAGCATCAAGGTG 1307

```
OY 1932 GCGGCTGTGGCATTGTGAATTGGGGCACTGGCTGTGCCCTGGGCCAGAAAGCCAGGGCT 1991
DB 1308 GCGGCTATGTGGCATTGTGAAGCTGGGGTACCCTGTGCTTTGGCCCGAAGCCAGAGT 1367
OY 1992 CTACACCAAGAGTCAAGTGAATTCGGGAGTGAATCTTCCAGGCCATAAAGACTCACTCCGA 2051
DB 1368 GTACACCAAGAGTCAAGTGAATTCGGGAGTGAATCTTCCAGGCCATAAAGACTCACTCCGA 1427
OY 2052 AGCCAGCGGCAATGTGACCCAGCTCTGA-CCGGTGGCTTCTGCTGGC----- 2099
DB 1428 AGCCAGTGGCAATGTGACTCAAGCCCTGATCCCGCTCATCTGCTGCTCCGCTGCACT 1487
OY 2100 AGCCTCCAGGGGCCGAGGTGAT-----CCCGTGGTGGATCCAGCTGG 2144
DB 1488 AGCATCCAGAGTCAAGAGTGTGTGGTGGCTCCAGCCCACTGTAGGCTCCACACTGG 1547
OY 2145 GCCGAGATGGGACGTTTTTCTTCTTGGGCCCCGCTCCACAGGTCCAGGACACCTCCCT 2204
DB 1548 GCCTCACAATGGAAATGTTCTCTGCTCAGATCCAGTCCAGGGTCCAGG--ATGCTGGAT 1605
OY 2205 CCAGGGTCTCTCTTCCACAGTGGGGGCCCACTGAGCCCCGAGACCAACCACTCACC 2264
DB 1606 CCAAGGACTTCTCTTCCACAGTGGGCCGCCCACTCAATCCAGGGCCATGGGCTCACC 1665
OY 2265 CTCTGACCCCCCATGTAAATATGTCTGCTGTCTGAGACTCTGTCTAGTGGCCCCCTGA 2324
DB 1666 TCCC--ACCCCATGTAAATATTAATCTGTCTCTGCGGGGGGCTCTAGGAGGCCCT-- 1720
OY 2325 TGATGGATGCTCTTAAATATAATAAAGATGTTTGATT 2363
DB 1721 TGTGCAAGTCTCTTTAAATATAATAAAGGTGGTTTGATT 1759
```

Search completed: August 17, 2005, 07:27:19
Job time : 16716 secs

this Page blank (uspto)